TITLE MOLECULAR CLONING OF THE MAJOR CASTOR BEAN 2S ALBUMIN PRECURSOR

AUTHOR STEPHEN D. IRWIN B.Sc.

INSTITUTION and DATE UNIVERSITY OF WARWICK 1989

MOLECULAR CLONING OF THE MAJOR CASTOR BEAN 2S ALBUMIN PRECURSOR

by

STEPHEN D. IRWIN B.Sc. (WARWICK)

A THESIS PRESENTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOLOGICAL SCIENCES
UNIVERSITY OF WARWICK

SEPTEMBER 1989

To Charlotte

CONTENTS

List of Figures

List of Tables

Acknowledgements

I acknowledge financial support from my parents and from the Science and Engineering Research Council.

The Plant Biochemistry group at the Department of Biological Sciences has provided me with an excellent training over the past three years. There are numerous members of the group I would wish to thank, both past and present. Particular thanks must be given to; Dr. Barry McGurl for his invaluable studies into the area and the donation of the 2S albumin antibodies; Jim Tregear (Ph.D. Pending) for the castor bean genomic DNA, library, and developing RNA samples; Saskia van der Vies for biological computing facilities; Drs. Andy and Mary Ryan for the napin antibodies, tricine gel methods, and a zealous interest in my work; Dr. Robert Spooner for the onerous task of teaching me molecular biology (and the short-cuts); and the lads and League of Spinsters who provided that 'je ne c'est quoi' that made me stick with it. Those not mentioned should not feel excluded.

Finally I would wish to thank my supervisor Professor Mike Lord, for providing me with everything I needed, including invaluable help, encouragement and ideas. Merci beaucoup, as they say in France.

Declaration

The data contained within this thesis are the results of original research conducted by the author under the supervision of Professor J.M. Lord. All experiments were performed by the author. Where advice or starting material was given, this has been acknowledged.

None of the data presented in this thesis have been used in a previous application for a degree.

Abbreviations

A	absorbance at the wavelength indicated as a subscript
amp	ampicillin
AMV	avian myeloblastosis virus
bp	base pairs
BSA	bovine serum albumin
Bq	Bequerel (also MBq, TBq)
C	centigrade
CaMV	cauliflower mosaic virus
CAT	chloramphenicol acetyl transferase
CIP	calf intestinal pancreas
cDNA	complementary DNA
CM	carboxymethyl
cm	centimetre (also um, nm)
Co.	company
CPM	counts per minute
CTF	CAAT-binding transcription factor
Da	Dalton (also kDa)
DEPC	diethyl pyrocarbonate
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DNPase	DNA polymerase
(d)(d)NTP	(di)(deoxy)nucleoside triphosphate (A = adenosine, C = cytidine, G = guanosine and T = thymidine)

DPM	disintegrations per minute
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediamine tetraacetic acid
ELISA	enzyme-linked immunosorbent assay
ER	endoplasmic reticulum
g	gramme (also kg, mg, ug, ng, pg)
GUS	beta-glucoronidase
h	hour
HEPES	N-2-hydroxyethylpiperazine-N-2-ethanesulphonic acid
HMW	high molecular weight
HPLC	high pressure liquid chromatography
HSTF	heat shock transcription factor
Ig	immunoglobulin
IPTG	isopropyl-B-D-thiogalactopyranoside
kb	kilobase
L	litre (also ml, ul)
LMW	low molecular weight
Ltd.	limited
M	Molar (also mM, uM, nM, pM)
mA	milliamps
2-ME	2-mercaptoethanol
min	minute
MOPS	4-morpholinepropanesulphonic acid
M_R	relative molecular mass
mRNA	messenger RNA
NET	NaCl/EDTA/tris

NF-1	nuclear factor 1
NOS	nopaline synthase
NT	nick translation
ORF	open reading frame
PBS	phosphate buffered saline
PHA	phytohaemagglutinin
plc	public limited company
poly A^+	RNA molecules that possess a polyadenylated 3' end
ppm	parts per million
RCA	*Ricinus communis* agglutinin
Rd.	road
rER	rough endoplasmic reticulum
RNA	ribonucleic acid
RNAse	ribonuclease
RNAsin	human placental ribonuclease inhibitor
RPM	revolutions per minute
RT	reverse transcriptase
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SSC	salt/sodium citrate
SSPE	salt/sodium phosphate/EDTA
TAE	tris/acetate/EDTA
TBE	tris/borate/EDTA
TCA	trichloroacetic acid
TE	tris/EDTA
TEMED	N,N,N,N'tetramethylenediamine

TEP	tris/EDTA/phosphate
Tris	tris (hydroxymethyl) aminomethane
tRNA	transfer RNA
U	unit
UK	United Kingdom
USA	United States of America
UTP	uridine triphosphate
UV	ultra-violet
v/v	volume to volume
Warwick	University of Warwick, Coventry
w/v	weight to volume
Xgal	5-bromo-4-chloro-3-indolyl-B-D-galactopyranoside

Amino acids, both free and as aminocyl residues comprising proteins, are represented by both standard one letter and three letter codes.
Inorganic salts are represented by standard abbreviations.

SUMMARY

Rabbit antibodies raised against the glutamine-rich castor bean 2S albumin, a heterodimer of 4kDa and 7kDa polypeptides, had previously been shown to immunoprecipitate a 34kDa polypeptide from the total products formed when castor bean endosperm mRNA was translated *in vitro*.

A cDNA library was constructed using maturing castor bean endosperm mRNA as template. Clones containing sequences complementary to albumin mRNA were isolated by hybridisation using as a probe a mixture of synthetic oligonucleotides representing sequences predicted for a peptide present in the 2S albumin large subunit. The nucleotide sequence contained an open reading frame encoding a preproprotein of 258 amino acid residues. The preproprotein included both polypeptides of the previously sequenced 2S albumin. In addition, this precursor included two further glutamine-rich sequences which, in terms of their size and conserved cysteine residues typically found in seed proteins of the 2S albumin superfamily, possibly represent the small and large polypeptide subunits of a second heterodimeric storage protein. A post-translational processing scheme has been proposed which would result in a single preproprotein generating two distinct heterodimeric 2S albumins. Studies on the protein bodies of castor bean endosperm cells suggested the existence of a second 2S albumin.

Genomic Southern blots indicated that at least 4 genes encode *Ricinus* 2S albumin precursors, and a Northern developmental analysis suggested that expression of such genes is temporally regulated. Genomic clones were isolated that represented the albumin genes and one of these was sequenced fully, in order to study the nature of the promoter control regions further. On comparison with the upstream regulatory regions of other 2S albumin genes, a region of high homology spanning 70 base pairs was found. This included a repetitive domain noted in many seed protein genes and implicated in storage protein promoter specificity. It was also located in ricin found likewise in castor bean endosperm cells.

CHAPTER 1

INTRODUCTION

This thesis describes studies undertaken on the seed storage proteins of the castor bean plant, Ricinus communis. One protein in particular, the 2S albumin, was analysed by cloning its precursor.

The introduction is divided into four sections, the first being a broad introduction to the nature of seeds and the storage reserves they contain, in particular their proteins. The aim of this section is to provide a basic appreciation of the range and diversity of interests found in this area. The second section concentrates on the castor bean plant and the proteins located within the seeds. This provides the entrance for the third section which highlights three areas of interest that the 2S albumin possesses and the reasons for its further study. Finally a brief introduction to the regulation of plant genes is given, with the emphasis firmly laid on promoter regions that generate expression specificity of seed storage protein genes.

The results obtained overlap a number of areas of widespread or intense research interest. The introduction in no way aims to be a definitive account of all the research performed to date in each category. What it does aim to provide is the context for the results contained herein, by describing common knowledge and research performed to date in relevant areas. Naturally, some of the observations described in the introduction were made after the onset of the project, and this should be borne in mind.

I1

SEED STORAGE PROTEINS

II.1 Seeds and their reserves

The seed represents a critical phase in the life history of the new plant. The success of a new individual in terms of its time and place of establishment and the energies it requires for proliferative growth are largely defined by the nature of the seed. Of particular importance are the major storage reserves of protein, starch and lipid laid down during development and maturation of the seed that sustain the young plant in the early stages of growth before it becomes an independent autotrophic organism able to photosynthesise.

II.1.a Seed development

The seed develops from the fertilised ovule. In angiosperms the seed is usually composed of: the embryo which results from the fertilisation of the egg cell in the embryo sac with one of the male pollen tube nuclei; the endosperm which arises from the fusion of the two polar nuclei in the embryo sac with the other pollen tube nucleus; the perisperm formed from the nucellus; and the testa, or seed coat, formed from one or both of the integuments around the ovule. Whilst all mature seeds possess an embryo and many are surrounded by a seed coat (although this may be rudimentary), the extent to which the endosperm or perisperm persists varies between species. In some cases the outermost structure is a pericarp, defining the whole as a fruit not a seed. In gymnosperm seeds there is no fusion between the male and polar nuclei, this then leads to the formation of the endosperm. The tissue is haploid and is the modified megagametophyte. In some plants such as the dandelion, seeds

are produced by non-sexual processes such as apomixis, from the diploid cells in the ovule (Bewley and Black, 1985).

II.1.b Storage reserves

Of all the food for human consumption, 70% comes directly from seeds. Eight species of cereal contribute over 50% of total world food calories. Seven species of legume make a smaller total contribution but are more significant in specific geographical areas. Human nutrition is supplemented by meat products derived from animals fed on seeds. Because of the significance of these figures much research has been performed on the storage reserves of those plants that are intensively farmed. Far less attention has been paid to other seeds that have low (perceived) agronomic value.

Amongst expected cellular biochemicals seeds contain enhanced levels of particular proteins, carbohydrates, oils and fats. Since storage proteins form the basis of study in this thesis they will be considered separately in more detail later.

Carbohydrates found in seeds are predominantly starch or hemicellulose, and to a lesser extent amyloids and raffinose oligosaccharides. Occasionally free sugars are stored such as in sugar cane. Starch is composed of two related forms of glucose polymer, amylose and amylopectin, and is deposited as subcellular bodies called starch grains. Hemicelluloses occur as the predominant carbohydrate in some endospermic legumes often laid down as very thick cell walls. They are

composed of mannose polymers, chiefly as mannans but also as the side chain substituted derivatives galactomannans and glucomannans.

Seed lipids are water insoluble esters of glycerol and fatty acids and are known either as fats or as oils if liquid above 20^{o}C. The majority are unsaturated fatty acids, the two major oils in oilseed crops being oleic and linoleic acid. The castor bean seeds however contain ricinoleic acid, which is a hydroxylated derivative of oleic acid. Of the less predominant saturated fatty acids, palmitic acid is the most common in seed oils. The triglyceride reserves are laid down in subcellular oil bodies. In the castor bean seed endosperm, these occupy a substantial volume of the cell. The structure of the oil body has recently been characterised in oilseed rape where it has been proposed that formation initially occurs with no proteinaceous membrane, this being formed later in development by insertion of hydrophobic 19 kDa polypeptide oleins directly into the oil body by ribosomes (Murphy et al, 1989).

I1.2 Storage proteins: Definition and classification

A seed storage protein can be defined as any protein which i) accumulates in the seed in sufficient quantities, ii) occurs only in the seed and iii) can be hydrolysed to release the constituent amino acids that are then used as a source of reduced nitrogen by the seedling during germination and early growth (Higgins, 1984; Spencer, 1984). Because of the last criterion they often contain a high proportion of the amides glutamine and/or asparagine.

Different attempts have been made to classify storage proteins and the most commonly used is that of Osborne (1924) based on the solubilities of the respective proteins. During extraction he obtained:

Albumins:	extracted in water
Globulins:	extracted in salt solutions
Prolamins:	extracted in aqueous alcohols
Glutelins:	extracted in alkali or acid solvents after prolamin removal

This remains a wall accepted set of classifications and has been subject from time to time to modifications with the onset of better biochemical analyses (Kreis et al, 1985b). Altschul et al (1966) has argued that the albumins are either residual enzymes or other metabolic proteins and are not storage proteins per se, whilst others using the castor bean as a model suggest that the albumins contain true reserve proteins (Youle and Huang, 1978ab; Li et al, 1977).

The major storage proteins of legumes and other dicotyledonous plants are globulins whilst those of the monocotyledonous plants are prolamins and glutelins (Higgins, 1984). There are of course exceptions to this such as in oats where the major storage proteins are globulins.

Kreis et al (1985b) suggested that seed storage proteins could be classified according to their function within the seed. They proposed four classes. Metabolic proteins would include most albumins and many globulins which are known to be enzymatic, for example Jackbean urease and maize sucrose synthetase. Structural proteins would include the ribosomal proteins, certain proteins of the ER and cell wall glycoproteins. This would not represent a major grouping. A third group would be those proteins that had no other defined function but to act as a storage of nitrogen, carbon and sulphur during seed development for germination. These true storage proteins would have a tighter definition than that suggested earlier and would include such requirements as being temporally accumulated in the latter stages of seed development; being directly involved in nitrogen nutrition as witnessed by their proportional fluctuation during nitrogen excess and deprivation; their subcellular localisation in discrete protein bodies; and the absence of any other function. The fourth category would include miscellaneous proteins that possessed a sedimentation coefficient of 2-3S. This last group of proteins is described in more detail later when considering the 2S albumins of the castor bean plant, but suffice to say they include a microheterogeneous group of proteins which possess similar sizes, high levels of glutamine and cysteine (Youle and Huang, 1981) and include such well characterised members as the 1.7S napin of Brassica napus

(Lonnerdahl and Janson, 1972; Crouch *et al*, 1983) and many alpha amylase/ trypsin inhibitors (whose members are well reviewed by Garcia-Olmedo *et al*, 1987).

II.3 History of storage protein research and future perspectives

II.3.a Nutritional improvement

The reserve proteins of seeds are a major source of nutrition for man and livestock. Because of this and their abundance they have been under formal investigation by protein chemists at least since 1747, when Becarri isolated gluten from wheat (Osborne, 1924). Research in the eighteenth and nineteenth centuries was aimed at discovering the relationships between plant and animal proteins and their composition. In the latter half of the nineteenth century it was demonstrated that amino acids were the components of the protein molecule. Early twentieth century research on seed storage proteins was dominated by the work of Osborne whose classic review of the vegetable proteins is still quoted by many researchers today (Osborne, 1924). Osborne proposed separation methodologies and suggested that except in closely related species, no two species contained proteins that could not be distinguished from one another chemically. In 1926, Sumner purified the protein Jackbean urease and demonstrated its catalytic properties, hence implying that enzymes were proteins. As a result research in seed storage proteins became less intense as more effort went into determining the biocatalytic nature of proteins (Dieckert and Dieckert, 1976).

Only recently has there been a revival of interest in storage proteins since the advent of molecular biology and the advance in the genetic engineering of crop plants which has made it possible to envisage the direct interference with the nutritional composition of seeds.

The nutritional problem with cereals and legumes is that they contain limited amounts of certain amino acids which are essential to man and monogastric animals. Most cereals are deficient in lysine and to a lesser extent threonine, whilst legumes are deficient in the sulphur amino acids methionine and cystaine. Some seed crops, notably rice, have low overall protein levels but with a better overall balance. Whilst a hundred years of breeding has increased seed yields, protein levels, and other desirable traits such as baking quality of wheat and malting quality of barley, only limited progress has been made in improving the nutritional quality (Mantell et al. 1985).

Research efforts have been directed towards cloning the genes for storage proteins and transforming foreign plants or reintroducing them allowing expression either by their own or a different promoter. A comprehensive review of this large research area is described by Weising et al (1988). The foreign gene product is either rich in essential amino acids or if it is itself nutritionally deficient has been manipulated by site-directed mutagenesis so that non-essential amino acids are converted to essential ones.

The expression of foreign genes in dicotyledonous plants has been made commonplace by the use of Agarobacterium transformation vectors and plant regeneration techniques. There are however problems with the manipulation of monocotyledonous plants which include the important cereal crops. They are resistant to Agarobacterium Ti vector transformation and recalcitrant to regeneration. Other methods of transformation have been devised which include direct gene transfer using electroporation, microinjection, liposome encapsulation,

protoplast fusion and using high velocity nucleic acid coated microprojectiles (Weising et al. 1988). Whilst transient gene expression has been detected in protoplasts, mature plants have only been reported to be generated from maize and rice (Rhodes et al. 1988; Abdullah et al. 1986).

There are other problems in the path of the genetic manipulation of crop nutrient content. It had been suggested that the presence of extra lysine residues in the storage proteins might reduce their hydrophobicity and interfere with the correct intracellular processing and targeting to protein bodies. This in turn would reduce the filling of the seed. Preliminary studies have shown though that zeins, the storage proteins of maize, can be engineered to contain extra lysine and tryptophan, and that this does not affect translation, signal cleavage, protein stability, aggregation and assembly into membrane bound protein body like structures when in vitro generated transcripts are microinjected into Xenopus oocytes (Wallace et al. 1988).

Storage proteins are usually encoded by multigene families, so that the introduction of one engineered gene may have little effect on the overall amino acid composition of the seed. If large amounts of engineered protein were produced, this might radically effect the amino acid pool in the cell, especially for the introduced amino acids, thereby affecting the production of other proteins.

It had been thought that correct developmental regulation of the genes would not be possible, thereby causing problems for the plant, but from many examples of storage protein gene transfer to foreign plants it is

found that expression is regulated in a similar fashion both with respect to time and tissue. This implies conserved cis- and trans-acting factors are present in different plant species (Weising et al, 1988).

It also seems that proteins can be correctly sent to the protein bodies of seeds as was seen for phaseolin, the major storage protein of Phaseolus vulgaris, when expressed in tobacco (Greenwood and Chrispeels, 1985a). When this protein was later modified by increasing the number of methionine residues from three to nine by insertion of a 45 bp synthetic duplex, the protein was found to be expressed in a temporal and tissue specific manner, with correct N-glycosylation and assembly into trimers. It was also located in the endoplasmic reticulum (ER) and Golgi apparatus secretion vesicles. However it was absent from the protein bodies having probably been degraded there. Acid hydrolases present may have recognised a wrongly formed structure (Hoffman et al, 1988). Clearly involvement in the nutritional improvement of plants by genetic engineering will require much further fine-tuning, since it involves many more biochemical criteria to be satisfied than just those for heterologous expression.

II.3.b Protein body targeting

Hartig isolated protein granules from several oilseeds using non-aqueous solvents and named them aleuron grains (Hartig, 1855, 1856). Pfeffer (1872) analysed seeds from many species and saw that they contained structures composed of proteins, organic phosphorous and metals. He termed these protein grains which are now commonly known as protein bodies. It was suggested that these developed in the plastids (Mottier,

1921) or vacuoles (Graham et al, 1962, Engleman, 1966). Altschul et al (1964) decided to name the vacuolar proteins of seeds aleurins. The aleurin concept was furthered by Dieckert and Dieckert (1976) who suggested a common process for angiosperm aleurin biosynthesis and sequestration. This led to the idea that the reserve proteins of seeds although appearing variable, were a limited set of homologous proteins produced along a common evolutionary theme in angiosperms. The hypothesis drawn by Dieckert and Dieckert (1985) was that storage proteins were allowed to be changed liberally as long as certain minimum restraints were met to allow sequestration to the rough endoplasmic reticulum (rER) -Golgi complex. Not only does this hypothesis allow for liberal genetic engineering of plant seed proteins for crop improvement, it also complements the findings described later that many storage proteins from both the monocotyledonous and dicotyledonous lines share common ancestral origins. The results of the modified phaseolin in tobacco seeds described earlier suggests that although liberally changed proteins may still be sequestered into the Golgi complex, insertion of nutritious amino acid residues which affect the minimum structural constraints may affect folding and make the protein susceptible to hydrolysis in the protein bodies.

The true role of the Golgi body in mediating the transport of storage proteins is not fully understood. Evidence reviewed by Chrispeels (1984) suggests that the majority of storage proteins are transported by the Golgi, yet the signals required for this sorting are as yet undefined. It has been known for some time that storage proteins are present first in the rER after signal cleavage where core glycosylation may take

place, and then in the Golgi apparatus, in which modification of oligosaccharides may or may not occur (Hurkman et al, 1981; Chrispeels, 1983; Bollini et al, 1983; Bassuner et al, 1983; Greenwood and Chrispeels, 1985b; reviewed Herman et al, 1986; Vitale and Bollini, 1986; Della-Cioppa et al, 1987).

Only recently have crucial experiments been performed with transgenic plants to begin to determine the prerequisites for correct sorting (reviewed Herman et al, 1989). Voelker et al (1989) showed that correct intracellular routing of common bean storage proteins phaseolin and phytohaemagglutinin (PHA) to the protein storage vacuoles occurred in transgenic tobacco seeds. More significantly site-directed mutagenesis was used to remove asparagine-linked glycosylation sites in PHA. This protein was also accumulated in the storage vacuoles indicating that glycans did not confer targeting information. These observations indicate that both PHA and phaseolin have the full information content in themselves to allow correct translocation through the endomembrane system for accumulation in protein storage vacuoles.

This sorting system is unlike that found for mammalian cell lysosomal hydrolases which are sorted by the trans-Golgi network according to the content of mannose-6-phosphate groups (Creek and Sly, 1984). If these glycans are not present the proteins are transported by default from the cells by the secretory pathway (reviewed by Kelly, 1985; Pfeffer and Rothman, 1987; Burgess and Kelly, 1987). In yeast cells targeting of acid hydrolases uses a second signal (other than the one for entry to the lumen of the ER) which is a polypeptide portion of the protein (Johnson et al, 1987, Valls et al, 1987). When phaseolin L subunit is

synthesised in yeast it is targeted to yeast vacuoles (Tague and Chrispeels, 1987) and can re-route yeast invertase from the secretory pathway to the vacuoles when combined as a fusion protein (Tague and Chrispeels, 1988).

Voelker et al (1989) also showed that zein, the non-vacuolar storage protein of maize accumulated in the storage protein vacuoles of tobacco indicating that deposition also occurs in some proteins that may lack obvious vacuolar targeting signals. Perhaps it may be that the way in which proteins are deposited in storage vacuoles is by a default pathway, those proteins not destined for this requiring targeting signals for other locations. This is considered unlikely since the default pathway for plant cells like animal cells has been shown to be extracellular secretion (Dorel et al, 1988).

11.4 Diversity and relatedness of storage proteins

Most studies on storage proteins have been performed on cereals and legumes which are considered here. In cereals the major storage proteins in maize, barley and sorghum are of the prolamin type, in wheat it is glutelin whereas in oats globulins predominate. Naming of specific proteins can be confusing and follows no fixed pattern. The prolamins and glutelins of major cereals are: Gliadin, glutenin (wheat); hordein, hordenin (barley); oryzin, oryzenin (rice) respectively, whilst the prolamins of maize, oats and sorghum are zeins, avenins and kafirins respectively.

Generally the major storage proteins of legumes are the globulins which account for up to 70% of total seed nitrogen. The globulins are further subdivided into two groups which differ in their sizes and sedimentation coefficients, namely the 7S vicilins and the 11S legumins. The lesser storage protein group is the 2S albumins. Exceptions occur, especially in some of the oilseeds such as castor bean where the 2S albumins are a major storage protein. Again nomenclature can be confusing and is species specific. For example, in soybean cotyledons the major protein is the legumin glycinin, the 7S vicilins are represented by the beta and gamma conglycinins whilst the 2S albumins are the alpha-conglycinins. The legumin type proteins include those such as legumin from Vicia faba, legumin from Pisum sativum, arachin from Arachis hypogaea, edestin from Cannabis sativa, cocosin from Cocos nucifera, oat globulin from Avena sativa and pumpkin seed globulin from Curcurbita sp. The vicilins are

known as vicilin from P.sativum, phaseolin from Ph.vulgaris, alpha-conarachin from Arachis hypogaea and concocosin from Cocos nucifera.

Storage proteins are generally oligomeric, that is made up of two or more subunits which can be linked by a combination of intermolecular disulphide, hydrogen, ionic and hydrophobic bonds. Often the exact native associations are unknown or poorly reported. In zeins there are major subunits of 22.5 and 19 kDa with a minor subunit of 13.5 kDa. Isoelectric focussing shows that these are composed of at least 15 polypeptides. Gliadin is separable into 4 major holoproteins made up of at least 46 different polypeptides. In legumes, pea legumins contain two different acidic subunits of 40 kDa and six different basic subunits of 20 kDa joined by disulphide bridges. Soybean vicilin beta-conglycinin has an average molecular weight of 160 kDa and is made up of 3 alpha subunits of 57 kDa and a beta subunit of 42 kDa. These may combine into six isomeric forms. Much of the variation reported appears to be derived from the presence of multigene families for storage proteins and many mature proteins are generated from similar sized/arranged precursors, (extensively reviewed by Fowden and Miflin 1984). Despite the variation seen between storage proteins there is a consensus of opinion which suggests that many storage proteins are evolutionarily related from a common ancestral gene. This is generated from comparisons of primary structure data generated by direct protein sequencing and derived sequences from cDNA and genomic clones. Prolamins in monocotyledonous plant seeds were thought to be a unique group because of their unusual amino acid composition and solubility properties, and so must have evolved recently in the ancestors of the modern grasses.

Figure 1

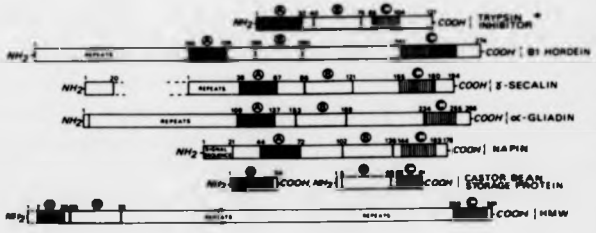
Schematic representations of the positions of the A, B and C regions in prolamins and related proteins.

* Similar regions of homology are present in wheat alpha-amylase inhibitor (Maeda et al. 1983) and in the millet bifunctional alpha-amylase/trypsin inhibitor (Kashlan and Richardson, 1981; Campos and Richardson, 1983). The sequences are numbered according to the original authors. The N- and C-termini of the proteins are indicated by 'NH_2' and 'COOH' respectively. 'HMW' is the high molecular weight prolamin of wheat.

With the exception of napin the residues are numbered from the N-terminus of the mature proteins. The residues of the napin are numbered from the start of the signal sequence because the site of signal cleavage is not known.

The sequences are taken from Odani et al. 1983; Forde B.G. et al. 1985; Kreis et al. 1985ab; Kasarda et al. 1984; Crouch et al. 1983; Sharief and Li, 1982; Forde J. et al. 1983. The sequences of the alpha-gliadin, trypsin inhibitor and castor bean storage protein were determined by direct sequencing of the protein. The others were deduced from the nucleotide sequences of cDNA or genomic DNA.

Reprinted from Kreis M., Shewry P.R., Forde B.G., Forde J. and Miflin B.J. (1985) "The Structure and Evolution of Seed Storage Proteins." Oxford Surveys of Plant Molecular Biology 2 253-317, by kind permission of the Oxford University Press.


NH2
COOH | TRYPSIN INHIBITOR *
REPEATS
COOH | B1 HORDEIN
COOH | γ-SECALIN
COOH | α-GLIADIN
COOH | NAPIN
COOH
COOH | CASTOR BEAN STORAGE PROTEIN
COOH | HMW

The primary sequence comparisons suggests that this is not so (Shewry *et al*, 1984ab; Kreis *et al*, 1985ab; Shewry, 1988). Whilst there are large numbers of repetitive domains which are rich in proline, unique to the group and generate the physicochemical characteristics of the prolamins, there are three non-repetitive domains named A, B and C which are also located in two other classes of very different proteins. These are the small 2S albumins found in dicotyledonous plants (for example castor bean, lupin, Brazil nut and oilseed rape), and the alpha-amylase and trypsin inhibitors from cereals. Figure 1 shows the areas in the proteins where the conserved domains lie. The repeat regions in the gamma-secalin and HMW protein are contracted otherwise the figure has been drawn to scale. Whilst the castor bean storage protein is shown as two separate subunits it has been speculated that this is produced as a single precursor molecule like that of the *Brassica napus* napin. The figure and legend have been exactly reproduced from the original article. The limits of the napin signal sequence have been defined by Ericson *et al* (1986).

What is particularly significant is that the non-repetitive domains contain homologous cysteine residues, conserved in similar positions, suggesting a structural homology rather than a strict sequence homology (Krebbers *et al*, 1988). This is especially true of the 2S albumins and as will be seen later in the results here obtained, the arrangement of cysteine residues is considered important in predicting whether a protein can be classified as a storage protein of the albumin, prolamin, protease inhibitor class.

This then suggests that the origin of the gene family that encodes these storage proteins predates the divergence of monocotyledonous and dicotyledonous plants. The repetitive elements found in the prolamins probably duplicated much more recently possibly within the ancestors of the Graminaceae.

I2.1 The castor oil plant

I2.1.a Agriculture

The castor bean is the seed of the castor oil plant Ricinus communis, the only species of the monotypic genus Ricinus, of the Euphorbiaceae family. Common varieties farmed are Baker, McNair, Cimarron and Hale. All grow from between 1-10m high, except Hale which is a dwarf, growing to a height of 0.1-0.15m and often used as the plant of choice in laboratory studies. They have large palmately lobed leaves of various shapes and colours ranging from dark green to reddish green depending on the variety. Plants have a bushy appearance with an erect stem and branches at the base of the plant although some cultivars appear more tree-like. Flowers can be up to 0.6m in length, the castor requires cross pollination.

The origin of the plant is unknown. It is believed to have originated in Africa and spread to the Middle East as a wild growing plant. Records show that castor oil was used in lamps in Pharaonic Egypt 6,000 years ago and was present in India and China some 3,000 years ago. Because of the high oil content of the beans (some cultivars yielding over 50% of their weight in oil) it is grown commercially in many tropical and subtropical areas. More castor is grown in South America than anywhere else, Brazil being responsible for almost half of the world harvest.

One of the reasons for its cultivation is its undemanding requirements. All that is required is a good supply of water in the early stages of

growth, for the first four months after planting until a vigorous root system has developed. Afterwards the plant is very drought resistant, requiring no additional water. The soil can be very poor indeed. As a result the plant is found growing wild in many areas where it is often regarded as a troublesome weed. Nevertheless the seeds of wild plants still contain a high oil content and it has been estimated that perhaps half of the Brazillian harvest is gathered from wild growing castor.

I2.1.b Industry

The castor oil is extracted from the seeds by mechanical grinding followed by expression or solvent extraction. The largest producers of seed, Brazil, India and the USSR crush their own produce, though there are strong crushing industries in Europe. Oil extraction generates a residue known as castor meal which is used after fine grinding as a supplement to mixed fertilizers. Whilst the meal is moderately rich in organic matter and minerals it is not used as an animal feed because of the presence of a number of toxins, including ricin and various allergenic components. Although these can be detoxified, the presence of bean shell debris renders the meal abrasive to the intestines of livestock. The economics of the processes required for animal feed suitability make it unfeasible.

The castor oil, when refined is an almost colourless liquid and with none of the toxicity associated with the meal. Because it uniquely contains almost 90% ricinoleic acid, it is distinguished from all other vegetable oils and fats and possesses remarkable physicochemical properties. These include miscibility with alcohol in every

concentration, high heat resistance, cold tolerance and burning without residue. It is used in the paints, resins, coatings, lubrication, pharmaceutical and cosmetic industries (United Nations Industrial Development Organisation, 1974).

I2.2 Biochemistry of castor beans

The fruit of the castor bean is globular, green or reddish before maturity with soft spines and though smaller, resembles the fruit of the chestnut tree. Seed maturity is normally reached 9 weeks after pollination. At this point the outer hull of the fruit has shrunk into a dry brown capsule which may split and shed the three seeds located within. These are oval in shape, possessing a brittle, black speckled maroon coat that surrounds the white, oily endosperm, and cotyledons.

Most biochemical analyses performed on the seed investigate the two physiological developments that the seed undergoes, maturation and germination. Since this thesis is primarily concerned with the deposition of storage proteins in the seed rather than their subsequent hydrolysis the maturation process will be described more thoroughly, with reference to the types of storage protein and their location.

I2.2.a Maturation

The endosperm cells are the site of synthesis and deposition of reserve materials which are utilised during seed germination. These reserves include the lipids, the major source of carbon during early post-germinative growth, and storage proteins and lectins which following hydrolysis during germination serve as a source of amino acids for further protein synthesis. During their biosynthesis between 10 and 35 days after pollination (Gifford *et al*, 1982), the storage proteins and lectins are packaged into and stored within single membrane bound

protein bodies, a feature shared by many other plant species (Lott, 1980). Storage proteins and the organelles which house them are very abundant in the mature seed, where they account for almost 95% of the total protein in endosperm tissue (Greenwood and Bewley, 1982). Ricinus protein bodies contain phytin globoids and a single large protein crystalloid within a soluble protein matrix (Tully and Beevers, 1976; Youle and Huang 1976). The organelles are 10-15 um in diameter. Isolation, subfractionation and electrophoretic analysis has shown that the protein bodies contain three major protein fractions. Over 50% of the organellar protein is 11S globulin which constitutes the insoluble crystalloid (although this has been estimated to be as high as 70-80% by Tully and Beevers, 1976; Gifford et al. 1982). The protein body matrix is a mixture of two water-soluble protein fractions, the 7S lectins which constitute approximately 6% (Olsnes et al. 1974; Tully and Beevers 1976; Youle and Huang, 1976) and the 2S albumins accounting for approximately 20% (Tully and Beevers 1976; Youle and Huang, 1976; Li et al. 1977; Sharief and Li, 1982; Gifford et al. 1982; McGurl, 1986).

The cotyledons in castor bean are not the main areas of storage protein deposition unlike the majority of dicotyledonous seeds (Bewley and Black, 1978). However protein bodies have been detected in the parenchyma of castor bean cotyledons (Greenwood, 1983). Analysis of these has shown that the proteins contained therein differ both quantitatively and qualitatively from those found in the endosperm. The soluble fraction does not contain the same lectins found in the endosperm, and contains no glycoproteins. The 11S crystalloids are represented by similar, yet different proteins and represent only 10% of

total cotyledon proteins. Deposition occurs after the endosperm proteins, between 35 and 45 days post-pollination. Yet during germination these proteins are mobilised before the endosperm proteins, 1-2 days after seed imbibition, suggesting a storage role for them, supporting the growth of the germinating seed prior to the mobilisation of the major storage proteins in the endosperm (Kermode et al. 1985).

It would seem then that the 11S crystalloids, 7S lectins and 2S albumins are the major proteins to be found in the castor bean seeds. Most research has been performed upon these, and in particular the 7S lectins, because they contain the toxic protein ricin. These groups will be discussed individually later.

I2.2.b Germination

In castor beans desiccation terminates the maturation phase, and this drying step is believed to be a developmental switch (Kermode and Bewley 1986, 1988). The switch from maturation to germination sees the onset of a different pattern of protein synthesis within the cotyledons and endosperm (Kermode and Bewley 1985ab, 1986; Kermode et al. 1985). Following imbibition, storage proteins are mobilised from the protein bodies of the endosperm. This occurs while the protein bodies merge to form a central vacuole. Mobilisation of the 2S albumins and 11S crystalloids occurs 2 days after imbibition and is complete within 4 days. The loss of protein is accompanied by an increase in three proteolytic enzymes, one carboxypeptidase and two sulphydryl dependent aminopeptidases. The 7S lectins are mobilised only slowly casting doubt as to their role as storage proteins (Gifford et al. 1983).

Glyoxysomes, a specialised form of peroxisome, are important structures in castor bean seed germination. They are small organelles, 0.5-1.0 um in diameter, bounded by a single membrane that encloses an amorphous matrix. These organelles play a vital role in gluconeogenesis from fats using the glyoxylate cycle and beta-oxidation pathway (Beevers, 1979). The glyoxysomes and their enzymic components are absent in the mature seeds and are synthesised de novo in a rapid and large-scale manner during early post-germinative growth. The nature of the glyoxysome and its comparison to the storage protein body in the castor bean seeds has been reviewed recently and so repetition has been avoided (Lord et al, 1989).

I2.3 The 11S crystalloids

As previously mentioned, Osborne (1924) described one class of storage proteins as the 11S globulins. This class can be further subdivided into the legumins if soluble in sodium dodecyl sulphate (SDS) or urea, and the crystalloids if also requiring salt for full dissolution (Gifford and Bewley, 1983). Whilst the legumins of plants have been well studied, in comparison the crystalloids have been largely ignored. In castor bean the 11S globulin is represented by crystalloid protein of molecular weight 330 kDa and is the most abundant of the storage proteins there (Gifford et al, 1982). Early studies on the crystalloid complex by Youle and Huang (1976) and Tully and Beevers (1976) suggested that they consisted of several proteins between 50-60 kDa, possibly with minor components as small as 17 kDa. On reduction, the major 60 kDa component yielded polypeptides of 32 kDa and 15.8 kDa suggesting that the unreduced crystalloids were composed of two small subunits and one large subunit.

Gifford et al (1982) showed that the crystalloids were a family of at least four proteins which reduced to two complementary groups after reduction with 2-mercaptoethanol. By cytochemical staining and SDS polyacrylamide gel electrophoresis (SDS-PAGE) they concluded that the crystalloids were produced concurrently. This was contradicted later by the work of Fukasawa et al (1988) who showed by in vivo labelling studies that the largest of the globulins was synthesised 20 days after anthesis whereas the smaller ones were produced after 30 days. A fuller

picture of the structure of the crystalloid complex and its components was given by Gifford and Bewley (1983). They suggested it was a hexamer structure of different heterodimers. Each heterodimer contains a larger, acidic polypeptide (30-40 kDa) linked covalently to a smaller, basic polypeptide (20-30 kDa) by a single disulphide bond. This acidic-basic heterodimer structure is common for many globulins (Matta et al. 1981; Gatehouse et al. 1980; Kitamura et al. 1976). Unlike other globulins though, the acidic and basic polypeptides of four of the castor bean crystalloid heterodimers can associate to form tetramer associations of 100 kDa (Gifford and Bewley, 1984b).

Labelling experiments have shown that the crystalloid heterodimers are made in precursor form (Gifford and Bewley, 1984a; Lord 1985a; Fukasawa et al. 1988). Antibodies raised towards the crystalloid complex specifically precipitated the constituent polypeptides from crude homogenates of labelled endosperm tissue. During pulse-chase experiments the antibodies immunoprecipitated proteins in the region 50-60 kDa from the endoplasmic reticulum (ER) and dense vesicles. Unlike the mature proteins these were insensitive to thiol reducing agents. These precursors were seen to be rapidly cleaved to generate the authentic heterodimer polypeptide components. These appeared in the crystalloids of protein bodies. Precursors in the ER fraction were processed in vitro into crystalloid components by the matrix fraction of protein bodies isolated from dry castor bean seeds. The variation in crystalloid protein subunits was attributed to a multigene family and not post-translational processing since the precursors also existed as a range of isoelectric variants.

Hence it would seem that a number of variant heterodimers from a gene family are made in precursor form on the rER, where they are transported to the protein bodies. Subsequent cleavage by an enzyme(s) generates the mature subunits which combine to form a holoprotein of six heterodimers.

12.4 The 7S lectins

Lectins are a group of proteins and glycoproteins which possess specific binding sites for particular sugars. They are very abundant in certain tissues and are widely distributed amongst plants (Lord, 1985c). The lectin fraction of the castor bean seed endosperm consists of two components, ricin and Ricinus communis agglutinin (RCA) (Olsnes and Pihl, 1982). Ricin was the first plant lectin to be isolated and characterised over a century ago (Stillmark, 1888) and a history of its research over the last hundred years has been reviewed recently (Franz, 1988). Interest in this molecule stems from the fact that it is one of the most potently cytotoxic compounds known. A dose of 35 ug by mouth may prove fatal. Severe gastro-intestinal and other symptoms have occurred, and even death, after ingestion of castor beans. Subcutaneous injection of as little as 1 ug/kg body weight can produce symptoms of poisoning that include congestion and diffuse haemorrhages in the lungs and abdominal viscera (Ordman, 1955; Osborne et al, 1905). The plant may produce it as a defence against grazing. Because of its toxicity attempts have been, and are being made to harness it in association with antibodies to construct anti-tumorous chemotherapeutic drugs. The creation of immunotoxins as 'magic bullets' was first envisaged by Ehrlich (1891) and their current limitations have been reviewed recently (Lord et al, 1988).

Ricin consists of two distinct N-glycosylated polypeptides (the A and B chains) joined by a single disulphide bond. The A chain (32 kDa) is a

highly specific ribosomal RNA N-glycosidase, which irreversibly inactivates 60S ribosomal subunits, thereby causing cell death upon entry to the cytoplasm (Endo et al. 1987). In many ways it resembles other ribosome inactivating proteins (Barbieri and Stirpe, 1982). The A chain alone is non-toxic when added to intact cells and requires the action of the associated B chain to allow it to enter the cytoplasm. The B chain (34 kDa) is a galactose specific lectin (Pappenheimer et al. 1974). The whole ricin molecule binds to the surface of cells via the B chain sugar binding sites interacting with exposed galactosyl residues of surface glycoproteins or glycolipids. The molecule is then endocytosed via coated pits and coated vesicles (van Deurs et al. 1985). Ricin A chain then separates from the B chain when crossing the membrane of an intracellular compartment, possibly from the trans-Golgi cisternae (van Deurs et al. 1986). Once into the cytoplasm cell death follows after catalytic inactivation of the ribosomes (reviewed Barbieri and Stirpe, 1982).

RCA has a molecular weight of 130 kDa and consists of two ricin-like heterodimers (Cawley and Houston, 1979), held together by non-covalent forces. Each heterodimer consists of a toxic A chain (32 kDa) disulphide linked to a galactose binding B chain (36 kDa).

The cDNA clones for both ricin and RCA have been obtained (Lamb, 1984; Roberts et al. 1985) and the A chains possess 93% homology whilst the B chains possess 84%. Genes for ricin have also been obtained (Halling et al. 1985; Tregear, 1989). Both ricin and RCA are produced as single preproprotein precursors. Both possess cleavable signal sequences removed upon entry into the lumen of the rER during translation by

membrane bound polysomes. Segregation is also accompanied by disulphide bond formation and core glycosylation. Heterogeneity in sizes of the proricins and proRCAs is caused by differences in glycosylation (Roberts and Lord, 1981b; Lord, 1985a). From the ER proricin and proRCA move to the Golgi apparatus where post-translational modification of the oligosaccharide side chains takes place including the addition of fucose (Lord, 1985b). Proricin is transported from the Golgi apparatus to the protein bodies by transporting vesicles, which fuse membranes with the protein bodies to discharge their contents. These contents also include proRCA, proglobulins and proalbumins (Lord, 1985a). All the propeptides are then cleaved by one or more specific endoproteases (Harley and Lord, 1985) to yield disulphide bond heterodimers. In both ricin and RCA a linker is removed between the A and B chains along with an N-terminal portion which has been speculated as being a protein body targeting sequence (Roberts et al, 1987).

12.5 The 2S albumins

The number of 2S albumins present in castor bean seeds has been reported to be either one or two, with differing calculations of molecular weight. Tully and Beevers (1976) identified two major albumins of molecular weight 10.3 kDa and 12.5 kDa that were not glycosylated and were present in the protein body matrix. These proteins were not reducible with 2-mercaptoethanol. Youle and Huang (1976) in a simultaneous study reported only one 14 kDa albumin protein. This did not degrade within 5 days of germination and was deemed not to be a storage protein. Two years later Youle and Huang (1978a) reported a 2S fraction from protein bodies that contained several proteins of molecular weight 12 kDa. These were rapidly degraded during germination and were hence classified as storage proteins, particularly in the light of their high amide content. They also calculated the albumin content to be approximately 40% of total seed protein. This is regarded as an overestimate, the true figure lying between 10-20%.

One of the major problems with these earlier studies on the 2S albumins was the poor resolution of the small proteins obtained by disc-gel electrophoresis. This probably led to the miscalculation of molecular weights as well. Gifford et al (1982) found two major albumin proteins in the protein bodies with molecular weights 12 kDa and 10.5 kDa using better electrophoretic methods. Gifford et al (1983) confirmed the role of the albumins as storage proteins by monitoring their hydrolysis

during germination in association with the appearance of specific peptidases.

Li and colleagues (1977) performed N-terminal sequence analysis on a purified 2S albumin of apparent molecular weight 5 kDa, noticing a high glutamine content and an unusual ultraviolet (UV) adsorption spectrum. The latter was suggested as being caused by the high content of disulphide bonds and a low number of aromatic amino acids. Similar low UV adsorbances were also noted in proteins such as plant protease inhibitors (Gennis and Contor, 1976). They also suggested that within the sequence of 22 residues was a region of homology to a sequence in wheat alpha$_2$-gliadin implicated in coeliac disease. This comparison is too small to be valid though.

The full sequence of the protein was later obtained by direct protein sequencing (Sharief and Li, 1982). The 5 kDa protein was found to be the small subunit of a disulphide linked heterodimer with subunits of actual molecular weight 7 kDa and 4 kDa. These migrated differently on gels though. The small subunit was composed of 34 amino acids, the large subunit containing 61 amino acids. Sequence heterogeneity was found in the large subunit where both leucine and serine were found at position 33. The intact protein contained a high amount of glutamine/glutamate and cysteine. Homologies were found unexpectedly with the sweet protein thaumatin 1 from the fruit Thaumatococcus daniellii, and more predictably with plant protease inhibitors. Their initial observations were that the protein was resistant to the action of trypsin, chymotrypsin and thermolysin, perhaps expected since the active site of trypsin inhibitory activity was found to be conserved in the large

subunit. The protein was later confirmed by McGurl (1986) as a serine protease inhibitor.

Youle and Huang (1981) studied the 2S albumin storage proteins from a wide variety of oilseeds and other plant seeds from 11 families. They noted that in each, the 2S albumins composed between 20-60% of total seed protein. These were characterised as storage proteins because of their high amide content. Moreover, it was noticed that they possessed a high content of cysteine unlike other storage proteins such as the globulins. They were thus described as having the unique role of providing a sulphur reserve for germination. The 2S albumins were thus defined by their low molecular weight, high solubility in water, high cysteine content and extremely high nitrogen content.

As other 2S albumin sequences were obtained by either direct protein sequencing or deduction from DNA sequences it became clear that all possessed homologies with the castor bean 2S albumin, which acted as it were as a model for 2S albumin primary structure. Homologies with the castor bean 2S albumin found included the barley trypsin inhibitor (Odani *et al*, 1983b); conglutin-delta from lupin (Lilley and Inglis, 1986); barley seed protein CMd (Halford *et al*, 1988); napin from *Brassica napus* (Ericson *et al*, 1986); sunflower albumin storage protein (Allen *et al*, 1987); pea seed albumin (Higgins *et al*, 1986); Brazil nut methionine-rich albumin (Altenbach *et al*, 1987) and the 2S albumins from *Arabidopsis thaliana* (Krebbers *et al*, 1988).

Crucially, it was recognised by Kreis, Shewry and colleagues that the prolamins of cereal seeds and the trypsin inhibitors of many seeds

showed homology to the growing list of 2S albumins sequences (Shewry et al. 1984; Krebs et al. 1985ab). This suggested the existence of a large superfamily of related proteins as described earlier. In comparing the castor bean 2S albumin sequence with the Arabidopsis thaliana 2S albumins and the Brazil nut albumin it was noted that the positions of cysteine residues were conserved. It was suggested that this indicated a structural framework for the albumins and that elsewhere in the sequence liberal changes could be made (Krebbers et al. 1988; Altenbach et al. 1987). All of the comparisons made to the castor bean 2S albumin showed this cysteine conservation.

Comparisons to the castor bean 2S albumin subunits were made with the derived precursor sequences for the albumins of Arabidopsis, Brazil nut, and Brassica napin. All three precursors were preproalbumins with predicted or demonstrated N-terminal signal sequences, linker peptides separating the subunits of which the small subunit was at the amino end, and which were sent to the protein bodies with the removal of a peptide between the signal sequence and the small subunit. This peptide might act as a protein body targeting peptide. It was suggested that the castor bean 2S albumin might also be produced as a preproalbumin, and be processed following a similar pattern (Altenbach et al. 1987; Ericson et al. 1986; Allen et al. 1987).

It would seem likely that this is so, since both the 7S lectins and 11S crystalloids in castor bean are produced in prepropeptide forms. Research on the 2S albumin has shown that a precursor polypeptide does exist and this will be discussed next. As regards the mature storage proteins of the castor bean endosperm, it is clear that a common theme

is involved, since all are composed of heterodimeric subunits. In the case of ricin and the albumin this is only one heterodimer, in RCA it is two and with the crystalloid complex it is six. The identity of the second castor bean 2S albumin is unknown.

I3 FEATURES OF THE CASTOR BEAN 2S ALBUMIN

I3.1 The 2S albumin is a serine protease inhibitor

Many seeds contain large amounts of protease inhibitors, and their major role appears to be a defence mechanism for plants against insects, fungi, bacteria and viruses (Ryan, 1973, 1981; Garcia-Olmedo et al, 1987).

It was previously stated that the characterised 2S albumin of castor bean was predicted to possess protease inhibitory activity (Li et al, 1977), and this was demonstrated by McGurl (1986) who showed it possessed inhibitory properties towards trypsin but not chymotrypsin. Approximately one mole of albumin inhibited one mole of protease.

Limited homology was seen between the albumin and a soybean member of the Bowman-Birk family of serine protease inhibitors (Sharief and Li, 1982). In spite of this, the albumin structure was found to more closely resemble those of the serine protease inhibitors found in cereals (Odani et al, 1983b) which were classified by Garcia-Olmedo et al (1987) as the cereal trypsin/alpha-amylase inhibitor family.

This family of proteins contains a diverse range of members, the structural relationships of which have only been recently made apparent. They represent a major part of the albumin fractions of seed endosperms and show homologies not only to 2S albumins from non-cereal seeds but also to the Kazal secretory trypsin inhibitor from bovine pancreas and alpha-amylases of bacterial origin, suggesting a superfamily of proteins

distributed beyond the plant kingdom (Odani *et al*, 1983c; Halford *et al*, 1988).

Odani *et al* (1983b) suggested that the castor bean 2S albumin was probably synthesised as a precursor and that this might have serine protease inhibitory activity since the reactive site was located between the small and large subunits. Whilst Garcia-Olmedo *et al* (1987) noted that the 2S albumins from castor bean, lupin, oilseed rape and Brazil nut had no reported inhibitory activity, clearly the mature castor bean does, and the active site is created between the two subunits (McGurl, 1986). It is expected that the napin has no such activity since no reactive site is conserved (Ericson *et al*, 1986).

The presence of such an abundant protease inhibitor suggests that not only is the castor bean 2S albumin a major storage protein, it probably has a major role to play in the protection of the seed against predation. It is expected that by inhibiting hydrolytic enzymes it can produce a toxic if not lethal action on feeding organisms. This effect was remarkably demonstrated by Hilder *et al* (1987) who used constitutive expression of cowpea trypsin inhibitor in tobacco to protect the plant from attack by the larvae of the tobacco budworm *Heliothis virescens*. The inhibitor had no effect on human digestion. This type of crop protection seems particularly promising since it avoids the use of chemical repellants, which are coming under increasing criticism for their effect on the environment. The castor bean 2S albumin would make a good candidate for such experiments using food crops, since not only would it possess protease inhibitory activity, it would also supplement the nutritient value of the crop at the same time. Little is known of

its effect on human digestion because of its intimate association with ricin. However as will be seen later, because of its implication in generating allergies it may not be such a good choice for further study.

I3.2 The 2S albumin is produced from a much larger precursor

It was speculated earlier that the castor bean 2S albumin might be synthesised as a precursor protein like those observed for the castor bean 7S lectins and the 11S crystalloids, as well as the 2S albumins from other plants. A precursor for the castor bean 2S albumin was identified by Lord and co-workers. Antibodies raised against a crude 2S albumin fraction precipitated a 34 kDa non-glycosylated polypeptide that was produced from a rabbit reticulocyte *in vitro* translation of developing castor bean seed poly A^+ RNA. This protein could be N-terminally processed by microsomes to a 32.5 kDa protein, suggesting the presence of a signal sequence. The protein was also shown to run as a 22 kDa polypeptide under non-reducing conditions. This migration was explained as a result of a large number of intrachain disulphide bonds maintaining the protein in a compact globular structure. This type of anomalous migratory property has been seen with other cystine rich 2S albumin precursors, notably in sunflower where the albumin migrates as a 19kDa polypeptide under reducing conditions, and as a 14 kDa polypeptide under non-reducing conditions. During *in vivo* labelled pulse-chase experiments it was observed that the precursor initially appeared in the ER, but disappeared from there with the subsequent accumulation of several mature albumin polypeptides in the protein body matrix. These were larger than the sizes of the subunits noted by Sharief and Li (1982), suggesting incomplete processing. It was concluded that the 34

Figure 2

Immunoprecipitation of the protein precursor to the castor bean 2S albumin storage protein

Poly A^+ RNA extracted from castor bean seeds of developmental stages D and E (Roberts and Lord, 1981a) was *in vitro* translated in a rabbit reticulocyte lysate system and immunoprecipitated with antibodies raised towards the previously identified castor bean 2S albumin. Samples were resolved on a 15% (w/v) denaturing polyacrylamide gel buffered with glycine under reducing conditions.

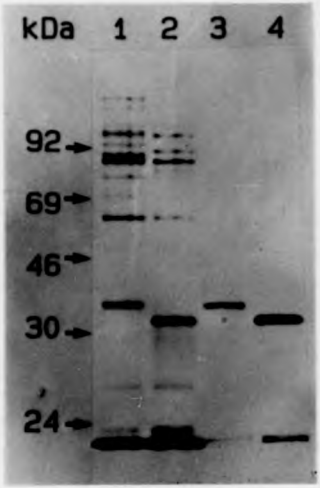
Lane 1 castor bean message translated in the absence of membranes.

Lane 2 castor bean message translated in the presence of membranes.

Lane 3 castor bean message translated in the absence of membranes and immunoprecipitated.

Lane 4 castor bean message translated in the presence of membranes and immunoprecipitated.

Membranes used were dog pancreatic microsomes. The gel was fluorographed, dried and exposed to film by autoradiography. Reproduced from Halpin C., Conder M.J. and Lord. J.M. (1989) with kind permission.


kDa
1
2
3
4
92
69
46
30
24

kDa protein was a 2S albumin precursor (Butterworth and Lord, 1983; Lord 1985a).

McGurl (1986) formally showed that the previously characterised 2S albumin (Li *et al*, 1977; Sharief and Li, 1982) was present in the 34 kDa precursor. Using purified albumin donated by Dr. S. Li, he generated rabbit polyclonal antibodies to the protein and used these to immunoprecipitate the 34 kDa protein from an *in vitro* translation of developing castor bean seed poly A^+ RNA. An example of the result is shown in Figure 2. McGurl managed to mimic the actions of the precursor that were observed earlier using the crude 2S albumin fraction antibodies. This included the cleavage of the signal sequence (Lanes 2 and 4). The figure also shows that the precursor is produced in abundance, suggesting high levels of expression during seed development as expected of a storage protein.

This precursor represents a preproalbumin -a signal sequence is removed cotranslationally and the proalbumin directed to the protein bodies where specific endoproteases generate the mature 2S albumin subunits. This type of enzyme activity has already been shown in castor bean protein bodies, and is not only able to to recognise the 7S lectins (Harley and Lord, 1985) and 11S crystalloids (Fukasawa *et al*, 1988) of castor bean as substrate, but also the proglobulins in pumpkin cotyledons, suggesting a conserved system of processing between plants (Hara-Nishimura and Nishimura, 1987).

There now appears to be a dilemma. If the maturing castor bean seed is actively synthesising storage proteins in large amounts to act as a

reserve for the progeny it would make sense for the plant to be efficient in the deposition process. Under evolutionary pressure, the plant should aim to donate as much reserve protein as possible to the seed to ensure survival of the plantlet before establishment. It seems unusual then that a 34 kDa precursor is produced in order to deposit a mature protein of molecular weight 11 kDa in the protein bodies. The 2S albumins of other plants exhibit larger precursor sizes than expected, but none of these are as seemingly wasteful as the castor bean. The napin of Brassica napus has a precursor of 21 kDa generating a mature protein of 13 kDa; in the Brazil nut sulphur rich albumin the reduction is from 18 kDa to 12 kDa; in radish albumin from 20 kDa to between 7 and 12 kDa; in sunflower albumin from 38 kDa to 19 kDa and for pea albumin from 13 kDa to 10 kDa (Crouch et al. 1983; Ericson et al. 1986; Altenbach et al. 1987; Laroche-Raynal and Delseny, 1986; Allen et al. 1987; and Higgins et al. 1986). In terms of sheer percentage loss to processing the castor bean 2S albumin is unique. Most storage proteins other than the 2S albumins do not share this high level of processing, only losing a signal peptide (Crouch et al. 1983).

McGurl (1986), in trying to assign a more reasonable use of such a large precursor, suggested that it may contain more than one copy of each previously characterised 2S albumin subunit or maybe more than one member of the 2S albumin fraction. The presence of multiple storage proteins on one precursor is rarely reported. The only examples include the two single chain albumins of pea which are neither disulphide-linked nor associated with each other in vivo (Higgins et al. 1986), and possibly the sunflower albumin precursor which has potential for a

second storage protein, although this has never been identified. Also, the mature sunflower albumin is unlike most albumins in that it is not a heterodimer but a single chain polypeptide. Hence, there are no reports to the author's knowledge of a single precursor generating two similar or dissimilar disulphide linked heterodimeric storage proteins.

I3.3 The 2S albumin is a potent allergen

I3.3.a Introduction to allergies

Castor beans give people allergies. In order to proceed with a description of this and the research performed so far it is necessary to possess a very basic understanding of the definitions and terms of the allergic process.

Allergy was first defined by von Pirquet in 1906 to describe all states of altered immunological reactivity in animals. This would include immunity to diseases as well as hypersensitivity. A more restricted view of allergy was adopted by Coombs and Gell (1975) who defined four states of hypersensitivity. Hypersensitivity can be best described as an inappropriate or exaggerated immune response which causes tissue damage. It is a characteristic of the individual and is manifested on second and subsequent contact with a particular antigen (thus known as an allergen). The first type of hypersensitivity, which is known as immediate, has become synonymous with the term allergy. This occurs when a particular type of antibody, immunoglobulin E (IgE) (Ishizaka and Ishizaka, 1967), is generated against innocuous allergens (such as pollen, faecal matter of house dust mites, insect stings, animal dander). This antibody becomes attached to circulating basophils and tissue mast cells which bear receptors for IgE. Upon second contact, cross-linking occurs resulting in the degranulation and release of vasoactive and pharmacological mediators such as histamine, bradykinin, and 5-hydroxytryptamine (Roitt et al. 1985). These generate symptoms

such as asthma, eczema, urticaria, and in severe cases anaphylaxis (vasodilation and constriction of smooth muscles, including those of the bronchus, which may result in death). Castor beans generate this immediate hypersensitivity. The other types of hypersensitivity include: type 2, cytotoxic; type 3, damage by antigen, antibody complexes; and type 4, delayed type or cell mediated hypersensitivity.

Allergens are named after the convention of Marsh et al (1988) after the organism from which the allergen originated. The first three letters refer to the generic name, followed by the first letter of the species name, followed by a Roman numeral (because of the existence of multiple allergens from individual species).

13.3.b Castor bean allergies

Meal generated from the extraction of castor oil from the seeds has been shown to cause severe hypersensitivity in those exposed to it. It is usually an occupational sensitiser affecting the people who handle the seeds and dust (Bernton, 1923, 1945; Robbins, 1923; Follweiler and Haley, 1925; Woringer, 1944; Garver, 1948; Ordman, 1955). These include oil-mill workers, dockers, fertiliser retailers and farmers. The latter two groups are affected because the meal is widely used as a fertiliser, which causes seasonal dust problems. Such is the potency of the castor bean allergens that hypersensitivity can be developed by those merely handling sacks that have contained castor beans (Snell, 1952). Indeed, there is a report of a case of a woman suffering anaphylaxis by wearing an Indian necklace of dried castor beans (Lockey and Dunkelberger, 1968). Generally speaking, all those who have contact with castor beans

can develop an allergy, including laboratory workers (Snell, 1924; Lord, Warwick, personal communication).

Epidemics of asthma have occurred mainly among citizens of towns containing castor oil mills, for example: Toledo, Ohio, USA (Figley and Elrod, 1928); Bauru, Brazil (Mendes and Ulhoa-Cintra, 1954); Johannesburg, South Africa (Ordman, 1955); Marseilles, France (Panzani, 1957); and more recently in Port Sudan (Kemeny et al, 1981). The problems in Marseilles and the Sudan have been well documented (Panzani et al, 1963; Kemeny et al, 1981). In recent years hygiene controls have reduced the frequency of sensitisation at least in some countries.

The dust from the meal is unusual in that it can sensitise people not predisposed to allergies. These individuals can produce precipitating IgE to castor bean extracts and develop symptoms of urticaria, conjunctivitis, rhinitis and asthma (Kemeny et al, 1981). Castor oil has no effect on castor bean sensitive patients (Bennett and Schwartz, 1934).

Identification of the castor bean allergens was initially undertaken decades ago by Spies and co-workers who concluded that they were a group of microheterogenous proteins of low molecular weight (Spies and Coulson, 1943). These proteins were present in a fraction named CB1-A and were water soluble, heat stable and possessed potent skin test activity in sensitised patients. Ricin had previously been eliminated as an allergenic agent (Barnard, 1930). Later Youle and Huang (1978b) showed that CB1-A and the 2S albumin storage proteins had similar mobilities on SDS polyacrylamide gels, and produced precipitating

antibodies when injected into rabbits. They concluded that CBI-A included the 2S albumins. Recently it was shown that three types of allergen occur in castor beans. The major allergen, named Ric cI, was identified in 96% of castor bean sensitive patients and was confirmed as the previously characterised 2S albumin. Ric cII was a crystalloid protein, and allergen 3 was a doublet of proteins of molecular weight 47 kDa and 51 kDa (McGurl, 1986; Thorpe et al. 1988). Whilst the albumins of pea (Higgins et al. 1986) and cottonseed (Youle and Huang 1979) have been implicated as allergens, neither give such a profound reaction as does the castor bean 2S albumin.

The previously characterised 2S albumin thus seems very interesting, for not only does it possess serine protease inhibitory activity, it is also produced as a precursor three times as large as itself, and is one of the most potent allergens known. Cloning of the 2S albumin precursor would be a useful step in the production of synthetic protein to be used in further studies of the nature of allergenicity. For example, because the allergen is unique in eliciting a response from non-predisposed individuals, site-directed mutagenesis could be performed to determine those sequences in the protein that may be required to elicit such a response. Already cDNAs have been obtained for the allergens Der pI, the major house dust mite allergen, and the white-faced hornet venom allergen Dol mV (Chua et al. 1988; Thomas et al. 1988, Fang et al. 1988). A sequence comparison between these might yield valuable information as to how they generate an IgE response in the first place.

I4 FACTORS AFFECTING PLANT GENE REGULATION

I4.1 Introduction to plant gene regulation

Genes are regulated in a highly controlled manner. They may be expressed in cells that have reached a particular developmental stage, or that are located in a particular set of tissues, or that have been stimulated by an environmental influence. Of course a combination of these events might be required. Modern molecular biology techniques now allow the unravelling of the biochemical and genetic factors that lie behind the complex system of gene regulation. What is becoming increasingly apparent is that whilst homologies can be found between plant and animal systems and, in a more conceptual fashion, between eukaryotes and prokaryotes, there is also a large degree of non-homology and incompatibility between systems. This has been especially seen during gene transfer experiments by many to induce protein synthesis in heterologous systems. Clearly any attempt to generalise and categorise a complex and incompletely known system is fraught with pitfalls and vulnerable to criticism. Nevertheless, an attempt will be made to describe the context of seed specific regulation of storage protein synthesis, in particular the regulatory sequences governing the expression of such. This is done so that an appreciation of the results in the thesis may be made later.

I4.2 Cis-and trans- acting factors in the regulation of genes

Production of proteins can be regulated in a variety of ways. These include: the regulation of transcription, by far the most common method; the regulation of translation of transcripts, for example haemoglobin biosynthesis in reticulocytes (Ochoa and deHaro, 1979), heat shock proteins (Altshuler and Mascarenhas, 1982) and oat globulin production during embryogenesis (Fabijanski and Altosaar, 1985); the regulation of the turnover of transcripts and protains; and the regulation of DNA replication, hence the number of genes being expressed, for example wheat seed endosperm cells during mid-stage development (Kowles and Phillips, 1985).

Attention is focussed here on the transcriptional control of gene regulation since this is believed to be the major controlling mechanism in the maturation of castor bean seed endosperm (Roberts and Lord, 1981a).

In higher eukaryotic cells, regulation of gene expression is often mediated by turning on or off RNA synthesis in a temporally ordered manner. Promoter selectivity requires the interaction of multiple cellular factors that recognise and bind to specific DNA sequences located within the promoter regions of eukaryotic genes. It is likely that both temporal regulation and tissue specific regulation of transcription are governed by these cellular DNA binding proteins (Dynan and Tjian, 1985; Briggs et al, 1986). A simplistic approach to describe the control of eukaryotic transcription would be to subdivide the

factors involved into DNA sequences located upstream (usually) of the transcription unit (cis-acting factors) and the soluble components that bind to these sequences thereby modulating the rate of transcription (trans-acting factors). The interaction of these cis- and trans- acting factors produces either a qualitative or quantitative effect, that is can increase or decrease the level of transcription, or switch on or off the gene. Unfortunately it is not often the case that the true role of a particular sequence and corresponding binding protein is known, since many operate in unison with a number of other modulating factors (Serfling et al, 1985). An example of this multiple control of expression is seen in the mammalian metallothionein I genes, where at least seven motifs have been recognised (Schibler and Sierra, 1987).

Studies on plant genes lie behind those of animal genes. Those plant promoters that have been analysed show similarities to animal genes. Plant scientists have thus been able to draw upon the knowledge amassed for the latter. Since the results in this thesis are primarily concerned with the cis-acting factors associated with the developmental and tissue specific transcriptional regulation of the castor bean 2S albumin gene, studies on trans-acting factors will be only briefly mentioned.

I4.2.a Trans-acting factors

These have begun to be recently isolated in animal systems by the techniques of gel retardation, DNAse footprinting and filter binding assays (Fried and Crothers, 1981; Galas and Schmitz, 1978; Diffley and Stillman, 1986), and their actions demonstrated in vivo or in vitro (Jones et al, 1987; Wiederrecht et al, 1987; Mizushima-Sugano and

Roeder, 1986; Scholer and Gruss, 1985; Briggs *et al*. 1986; Bazett-Jones *et al*. 1985; Lee *et al*. 1987).

Some of these factors have been found to be very similar, for example CAAT-binding transcription factor (CTF) and nuclear factor 1 (NF-1) both act as transcript selectivity factors for RNA polymerase II and as initiation factors for adenovirus DNA replication (Jones *et al*. 1987), and heat shock transcription factor (HSTF) from *Saccharomyces cerevisiae* can activate elements in other distantly related organisms such as *Drosophila melanogaster* (Wiederrecht *et al*. 1987). That the trans-acting factors are near identical in the latter example suggest strong conservation of this system in eukaryotic organisms. That foreign genes can be expressed in transformed organisms and be regulated in a native fashion also confirms this. For example, many storage protein genes are correctly regulated in a tissue and temporal manner in foreign plants. Abscisic acid has been seen to be involved with the expression of specific genes such as the a' subunit of beta-conglycinin from soybean and hence might indirectly influence a trans-acting factor in this process (Bray and Beachy, 1985; reviewed Quatrano, 1986). Undoubtedly more DNA binding proteins will be isolated and characterised providing more information on their specific action in gene regulation, since this is as yet poorly understood.

14.2.b Cis-acting factors

These DNA sequences have been characterised in a number of ways including comparisons with related genes and databank searches (Dickinson et al, 1988), and transgenic studies combining point mutation analysis (Grosschedl et al, 1981; Myers et al, 1986) and expression of reporter genes such as nopaline synthase (NOS), beta-glucoronidase (GUS) and chloramphenicol acetyl transferase (CAT) fused to the promoter regions of interest. These are linked if necessary to a controlling constitutive promoter such as the 35S of cauliflower mosaic virus (CaMV) (reviewed Weising et al, 1988). Whilst the first technique often helps define putative controlling sequences, the latter two techniques are necessary to validate the predictions.

Some elements have been found to be almost universally present in eukaryotic genes, and need to be present if transcription is to occur. These donate the quantitative effect on expression. Other elements appear to be more specific to certain types of gene and are more qualitative in their effects.

I4.3 Elements exerting a quantitative effect

I4.3.a The TATA box

Usually found as the consensus TATA(A/T)A in animal sequences it is often known as the Goldberg-Hogness box. In plants it is found to be similar in both monocotyledonous and dicotyledonous plants (Messing et al, 1983). Forde et al (1985) noted a longer consensus in monocotyledonous plants CTATA(T/A)A(T/A)A. The box appears to be necessary for the generation of faithful 5' termini of mRNAs (Breathnach and Chambon, 1981). Mutation of this region generates 5' heterogeneity in the transcripts (Grosschedl and Birnstiel, 1980; Benoist and Chambon, 1981; Ghosh et al, 1981). In both plant and animal genes it occurs approximately 25 bp upstream from the startpoint of transcription (the cap site). It is believed that the TATA sequence acts as a recognition site for RNA polymerase II which transcribes nuclear genes thereby generating mRNA.

I4.3.b The CAAT box

Approximately 80-100bp upstream of the cap site is another consensus sequence GG(C/T)CAATCT found in nearly all animal genes. It has been demonstrated important to transcription by the mutational analysis of beta-globin genes (Myers et al, 1986). In plants it is not so well conserved. Many genes do not possess this box. In some, a different box is found with consensus CATC. In zein and other genes a different consensus is seen in this region $(C/T)A_{2-5}(G/T)NGA_{2-4}(C/T)(C/T)$ and has

been called an AGGA box (Messing et al, 1983). The significance of this box has not been tested. The binding protein for the CAAT element has been identified as CTF (Jones et al, 1987).

I4.3.c Enhancer-like elements

These elements have been the subject of much controversy, particularly over their classification. Whilst it is known that enhancers possess properties that upstream regulatory elements do not (for example, long range influence) it is becoming increasingly clear that the two overlap both physically and functionally. With the characterisation of eukaryotic promoters as being composed of multiple modules, the strict distinction between enhancer and promoter element has been lost.

Enhancers isolated so far, mostly from animal and viral genes, possess the following properties:

1) strong activation of transcription of the linked gene from the correct cap site
2) orientation independence
3) function over long distances (greater than 1 kb) upstream or downstream of the cap site (Dynan and Tjian, 1985)

Most studies on enhancers have been performed on those found in SV40 early genes (Benoist and Chambon, 1981), immunoglobulin genes (Banerji et al, 1983), and metallothionein (Karin et al, 1984). Whilst enhancer sequences can be quite extensive, 72 bp in SV40 and 450 bp in cytomegalovirus, internal repeats often occur, containing what is known as the enhancer core GTGG(A/T)(A/T)(A/T)G (Weiher et al, 1983). Also

implicated in the structure of enhancers are inverted repeats such as the 19 bp repeat of consensus CCCATTGAC|GTCAATGGG in human cytomegalovirus early gene (Boshart et al. 1985) and the presence of alternating purine/pyrimidine groups known as RY repeats. However, different enhancer elements bear little homology. Most cellular enhancers and also some viral enhancers exhibit their activity in a highly cell-type specific manner (Sodroski et al. 1984; Mosthaf et al. 1985) making their categorisation as both quantitative and qualitative. This type of specificity is probably dependent on transacting factors in certain cells and tissues. It has been speculated that the overall function of these motifs is to bind various factors (Scholer and Gruss, 1984, 1985) which as a result make the DNA in the locality open to access by RNA polymerase II molecules (reviewed Serfling et al. 1985; Dynan and Tjian, 1985; Gruss, 1984; Wasylyk, 1988; Muller et al. 1988).

Enhancer-like elements have been noted in plant genes also. Not surprisingly, the first to be characterised was from a plant virus. The CaMV 35S promoter contains a 59 bp region that satisfies the criteria detailed above. Located within this region is the CAAT box and an inverted repeat containing the enhancer core sequence. Reports of enhancers in plant seed storage protein genes include pea legumins (Lycett et al. 1985), zeins (Roussell et al. 1988), and the soybean beta-conglycinin (Chen et al. 1986). As will be seen later, elements implicated in seed storage protein gene regulation have properties similar to enhancers, including RY repeats, the core sequence and inverted repeats.

I4.4 Elements exerting a qualitative effect

Distinction between the quantitative and qualitative elements can be difficult to justify when studies show that removal of the latter can generate a highly detrimental effect on gene transcription. Nevertheless, they are not ubiquitous and so are considered as a separate group. They can be subdivided into two overlapping groups, those that bind tissue or developmental stage specific effector molecules, and include members such as the seed storage protein gene elements, and those that can be considered inducible regulatory elements (Schibler and Sierra, 1987).

I4.4.a Inducible elements

Gene regulation of the inducible type has been seen in plants under various controlling influences. Some genes are regulated by interactions with microbes, such as the nodulin genes, for example leghaemoglobin (Hyldig-Nielsen et al. 1982) and defence genes such as the enzymes involved in lignin and chitin biosynthesis (Bell et al. 1986; Broglie et al. 1986). Others are known to be heat shock related, the products of which are similar to those found in animals, in particular Drosophila (Schoffl et al. 1986). They can be induced by a wide range of other stimuli and are expressed in most plant tissues (Cooper et al. 1984).

Possibly most research has been performed on the light induction of plant genes, and in particular for the ribulose bisphosphate carboxylase small subunit, rbc-S, and the chlorophyll a/b binding protein, cab

(reviewed Tobin and Silverthorne, 1985; Kuhlemeier et al. 1987). The pea rbcS-3A gene contains an enhancer-like element which contains the information necessary for light-regulated expression and cell-type specific expression (Aoyagi et al. 1988). Recently three cis-acting sequences necessary for U/V light activation of the parsley chalcone synthase gene were elucidated. One of these sequences was found to be homologous to an element implicated in the regulation of genes involved in photosynthesis (Schulze-Lefert et al. 1989). Clearly the light regulated gene elements overlap with the tissue-specific/developmental category.

I4.4.b Tissue specific and developmental elements

These studies in plants have mostly been performed upon seed storage proteins. This is because they are abundant, hence easier to clone, are tightly regulated and are specific to defined tissues within the seed. Some of the literature involved in identifying the cis-acting sequences for temporal and tissue specific regulation is located in the Discussion section (D4.6) to avoid too much cross-referencing between chapters. From the sequences that are emerging as important in storage protein regulation, by homology studies and deletion/transformation experiments, it seems that two major boxes can be located, the cereal box and the legumin box. These are not mutually exclusive, since the cereal box has also been found in pea legumin genes (Lycett et al. 1985) and the legumin box can be noted in prolamins (Forde et al. 1985). It just so happens that the reports of the two boxes appear to be roughly split between the monocotyledonous and dicotyledonous lines.

The cereal box

This was noted in the prolamins of barley, wheat and rye by Forde et al (1985) as a conserved region 300 bp upstream of the translation start point. It was of the consensus TGTAAAG. That this was similar to the core consensus sequence of the SV40 enhancer was not noted. Simultaneously, Lycett et al (1985) noted the SV40 core enhancer sequence in three legumin genes. This immediately led to the speculation that they possibly represented sequences important in specific transcriptional activation. Other plant genes containing this sequence emerged (Baumlein et al, 1986; Doyle et al, 1986; Timko et al, 1985; Kaulen et al, 1986). Chen et al (1986) noted however that addition of an 8 kb fragment containing the box had no significant effect on the transcription of a' subunit of soybean beta-conglycinin in transgenic Petunia plants. Colot et al (1987) showed that a region between 160 bp and 326 bp upstream of the cap site of the wheat low molecular weight (LMW) glutelin gene was sufficient for endosperm specific expression of CAT in transgenic tobacco. This region included two copies of the cereal box. Roussell et al (1988) made a similar observation but suggested 5 regions possessing homologies to the SV40 enhancer core/cereal box were present. Some of the most conclusive evidence for the implication of this element in specific transcription activation came from Maier et al (1987) who showed that crude nuclear extracts from maize endosperm recognised a 22 bp binding site located in the 5' flanking region of a zein gene. The binding site included 14 bp of a 15 bp sequence conserved in zein genes. This site also contained the cereal box.

The legumin box

Hoffman and Donaldson (1985) noted the presence of alternating pyrimidines and purines in the 5' flanking regions of two Phaseolus vulgaris phytohaemagglutinin (PHA) genes. These genes were found to be expressed specifically in cotyledon tissue. One set of RY repeats closely flanked the cap site whilst the other four were present up to 230 bp upstream of the translation start point. These repeats included the common sequence ATGCAT. The motif was noted to be present in Phaseolus lectins obtained previously (Hoffman, 1984; Vodkin et al. 1983). Since RY repeats had previously been seen to promote the formation of Z-DNA (Nordheim and Rich, 1983) and that DNA was seen to be sensitive to the action of S1 nuclease at the junction of RY repeats with non-RY DNA (Kilpatrick et al. 1984) it was suggested that the formation of Z-DNA near the cap site might improve the accessibility of the transcription initiation region to RNA polymerase II. The enhancer core/cereal box was absent.

Baumlein et al (1986) showed a region between -80 and -107 from the cap site of a legumin gene from Vicia faba was highly conserved between legumin genes including those from soybean and pea. The region was tentatively named a "legumin box" and was suggested to be a regulatory element specific for this group of seed protein genes. Gatehouse et al (1988) noted that the putative enhancer sequence identified by Lycett et al (1985), the cereal box, was not strongly conserved in two genes encoding pea 'minor' legumin polypeptides. Strong conservation however was seen with the "legumin" box. They concluded that the latter region

was likely to be involved in determining the high tissue specific expression of these genes. This claim remains to be substantiated.

Dickinson et al (1988) compared the 5' flanking regions of 21 legume seed protein genes, including lectin genes, legumin type genes, vicilin type genes and one seed albumin gene. Of these 20 contained at least 7 bp of homology to the 8 bp RY sequence CATGCATG. Some genes possessed multiple copies of the element, all were present within 290 bp of the cap site. They also noted the presence of the motif in many cereal seed protein genes at or near the -300 element, the cereal box, described by Forde et al (1985). The CATGCATG motif was absent from 31 non-seed genes when analysed up to 300 bp upstream of their respective cap sites. Dickinson et al (1988) concluded that the motif played a role in the regulation of transcription of these genes, probably manifested as some means of developmental or tissue specific influence. The motif was present within the "legumin boxes" identified.

In their studies of the Brassica napus napin, Josefsson et al (1987) and Scofield and Crouch (1987) noted the presence of RY repeats in the 5' upstream regions. Krebbers et al (1988) noted the conservation of the CATGCATG motif in albumin genes from Arabidopsis thaliana.

It would seem then that whilst the longer "legumin box" of Baumlein, Gatehouse and colleagues may be restricted to legumins, it is clear that the shorter motif CATGCAT is not.

CHAPTER 2

MATERIALS AND METHODS

CHAPTER 2a

MATERIALS

Amersham International plc, Amersham, Buckinghamshire, UK: Radiochemicals; biotin-streptavidin Western blotting system; Hybond-C nitrocellulose mebranes; Hybond-N nylon membranes; restriction enzymes; T4 DNA ligase; T4 polynucleotide kinase; T4 DNA polymerase; E.coli polymerase 1 (Klenow fragment); lambda DNA; Eco RI linkers; Amplify; dog pancreatic microsomal membranes.

Anderman and Co. Ltd., London Road, Kingston-upon-Thames, Surrey, UK: Schleicher and Schuell Biotrap membranes.

BDH Chemicals Ltd., Poole, Dorset, UK: acrylamide; ammonium acetate; ammonium persulphate; bromophenol blue; citric acid; 2-mercaptoethanol; phenol; sucrose; potassium chloride; sodium hydroxide; N,N,N,N'-tetramethylenediamine (TEMED); zinc chloride.

Beckman-RIIC Ltd., High Wycombe, UK: non-aqueous scintillant; Quickseal tubes;

Bio-rad Laboratories Ltd., Watford, Hertfordshire, UK: Biogel P-60.

Boehringer Mannheim (BCL), Lewes, East Sussex, UK: deoxy- and dideoxynucleotides; E.coli tRNA; ribonucleotides; E.coli DNA polymerase 1 (nuclease free); calf intestinal alkaline phosphatase; creatine kinase; creatine phosphate; oligo-dT_{12-18}.

Calbiochem, La Jolla, California, USA: Aquacide.

Clause (UK) Ltd., Charvil, UK: castor bean seeds.

Difco Laboratories, Detroit, Michigan, USA: Bacto-tryptone; yeast extract; agar; bacto-gelatin.

Eastman Kodak, Rochester, New York, USA: N,N'-Methylene bisacrylamide; photographic film (Panatomic-X, Plus X-Pan); paper (Kodabrome II)

Fisons Scientific Apparatus, Loughborough, Leicestershire, UK: dimethyl sulphoxide; ammonia; formaldehyde; glycine; hydrogen peroxide; urea; glycerol; sodium chloride; diaminoethanetetra-acetic acid (EDTA); magnesium chloride; magnesium sulphate; calcium chloride; sodium dodecyl sulphate; isobutanol; isopropanol; methanol; ethanol; caesium chloride; boric acid; ammonium sulphate; sodium dihydrogen phosphate; disodium hydrogen phosphate; sodium acetate; magnesium acetate; potassium acetate.

Fuji Photo Film Co. Ltd., Japan: X-ray film (RX).

Gibco-BRL Ltd., Uxbridge, Middlesex, UK: Restriction enzymes; oligo-dT cellulose; sharkstooth combs.

Ilford Ltd., Mobberley, Cheshire: FF Contrast developer.

Imperial Chemical Industries (Pharmaceuticals Division), Macclesfield, UK: Mixed synthetic oligonucleotide.

Life Sciences Inc., North Saint Petersburg, Florida, USA: Avian myeloblastosis virus reverse transcriptase.

May and Baker Ltd., Dagenham, Kent, UK: Hydrochloric acid; glacial acetic acid; trichloroacetic acid; chloroform; diethyl ether.

New England Biolabs, Beverley, Massachusetts, USA: Eco R1 Methylase; S-adenosyl-methionine.

P and S Biochemicals Ltd., Liverpool, UK: RNAsin.

Pharmacia (GB) Ltd., London, UK: Sephadex G-50; Sephadex G-15; DEAE-Sephacel; protein A Sepharose; M13 universal primer; E.coli DNA polymerase 1 (Klenow fragment) FPLC pure; T7 RNA polymerase; 5' cap; hexadeoxyribonucleotides.

Polaroid Corporation, Cambridge, Massachusetts, USA: Polaroid 667 film

Promega-Biotech, Madison, Wisconsin, USA: pGEM3-blue.

Sigma Chemical Co. Ltd., Poole, Dorset, UK: Dithiothreitol; Coomassie brilliant blue R; silver nitrate; bovine serum albumin; molecular weight markers; Triton X-100; 3'-diaminobenzidine; Tris-Cl base; Nonidet P-40; imidazole; S1 nuclease; lysozyme; ethidium bromide; ficoll 4000; bromocresol green; spermidine; spermine; ampicillin; 5-bromo-4-chloro-3-indolyl-B-D-galactopyranoside; isopropyl-B-D-thiogalactopyranoside; poly-A DNA; xylene cyanol; polyethylene glycol; NZ-amine; trinitriloacetic acid; Tween-20; diethyl pyrocarbonate (DEPC); DNAse; 4-(2-hydroxyethyl)-1-piperazineethanesulphonic acid (HEPES); 4-morpholinepropanesulphonic acid (MOPS); agarose; tricine; amino acids; rubidium chloride; manganous chloride; polyvinylpyrollidone.

United States Biochemical Corp., Cleveland, Ohio, USA: Sequenase kit

Whatman Labsales Ltd., Maidstone, Kent, UK: filter paper (3MM, No.1 and GF/C); CM-cellulose.

CHAPTER 2b

METHODS

M1.1 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

M1.1.a. Glycine buffered system.

The discontinuous buffer system of Laemmli (1970) was used with gels measuring 22 x 20 x 0.15 cm, whenever protains of molecular weight greater than 14 kDa were to be analysed. Resolving gels were made as follows, depending on the final concentration of acrylamide required. Stock acrylamide contained: acrylamide (30% w/v) and bis-acrylamide (0.8% w/v).

Resolving gel mix

Final [acrylamide]	10%	12%	15%	20%
Acrylamide stock	16	20	24	32
3M Tris-HCl (pH8.8)	6	6	6	6
H_2O	25.4	21.4	17.4	9.4
SDS (10% w/v)	0.48	0.48	0.48	0.48

Polymerisation proceeded on the addition of 200ul ammonium persulphate (10% w/v) and 20ul TEMED. Water saturated isobutanol was layered onto the polymerising resolving gel. Once set, the isobutanol was removed by thorough washing with distilled water and replaced with polymerising stacking gel mixture, followed by gel comb insertion. Stacking

acrylamide contained: acrylamide (10% w/v) and bis-acrylamide (0.8% w/v).

Stacking gel mix	mls
Stacking acrylamide	5
0.5M Tris-HCl (pH 6.8)	2.4
H_2O	2.4
SDS (10% w/v)	0.1

Polymerisation proceeded on the addition of 100ul ammonium persulphate (10% w/v) and 5ul TEMED. Once set, the comb was removed and the gel placed into a vertical electrophoresis tank with buffer in both top and bottom reservoirs containing 25 mM Tris-HCl (pH 8.8), 192 mM glycine and SDS (0.1% w/v).

Samples were mixed with loading buffer and boiled for five minutes. The type of loading buffer used depended on the analysis sought. Under non-reducing conditions, 2-mercaptoethanol was replaced by water in the loading buffer.

Loading buffer (2x concentration)	mls
0.5M Tris-HCl (pH 6.8)	2.5
Glycerol	2
SDS (10% w/v)	4
2-mercaptoethanol	1
H_2O	0.5
Bromophenol blue	1mg

Molecular weight marker protains were run routinely and consisted of:

Protein	kDa
Phosphorylase b	94
BSA	67
Ovalbumin	43
Carbonic anhydrase	30
Soybean trypsin inhibitor	20.1
a-Lactalbumin	14.4

Where gels were run for subsequent Western blotting, prestained molecular weight markers were included and consisted of the following proteins:

Protein	kDa
a-Macroglobulin	180
b-Galactosidase	116
Fructose-6-phosphate kinase	84
Pyruvate kinase	58
Fumarase	48.5
Lactate dehydrogenase	36.5
Triosephosphate isomerase	26.6

Gels were typically run with constant current at 10 mA overnight or 25 mA during the day.

M1.1.b Tricine buffered system.

This system, adapted from Shagger and von Jagow (1987) by Dr. A. Ryan (personal communication) was used typically to analyse the 2S albumin

fraction of castor beans and enabled the separation of proteins in the range from 1 to 100 kDa. The major difference between this system and that of the glycine system of Laemmli (1970) occurs in the reservoir buffer where the analogue tricine replaces glycine. The gel mixes used are described below [note no spacer gel as described in Shagger and von Jagow (1987) was used]. Gels were thinner, 22 x 20 x 0.08 cm and took noticeably longer to run than glycine gels, typically 36-48 hours. Stock acrylamide contained: acrylamide (48% w/v), bis-acrylamide (1.5% w/v).

Resolving gel mix	mls
stock acrylamide	10
3M Tris-HCl (pH 8.45)/ SDS (0.3% w/v)	10
H_2O	9.9

This yielded a final acrylamide concentration of 16.5%. Polymerisation proceeded as usual after the addition of 100 ul ammonium persulphate (10% w/v) and 10 ul TEMED. Stacker gel mix was then layered over the polymerised gel as usual.

Stacker gel mix	mls
stock acrylamide	1
3M Tris-HCl (pH 8.45)/ SDS (0.3% w/v)	3.1
H_2O	8.3

Polymerisation occurred with the gel comb in place after the addition of 100 ul ammonium persulphate (10% w/v) and 10 ul TEMED. Protein samples were prepared in the same way as before using loading buffer that contained 4M urea.

Loading buffer (2x concentration)

SDS	4% (w/v)
Glycerol	12% (v/v)
Tris-HCl (pH 6.8)	50mM
2-mercaptoethanol	2% (v/v)
Urea	4M
Bromophenol blue	0.01% (w/v)

Marker proteins were different in this system, reflecting the sizes of proteins under investigation.

Protein	kDa
Myoglobin (polypeptide backbone)	16.95
Myoglobin (fragment I+II)	14.4
Myoglobin (fragment I)	8.16
Myoglobin (fragment II)	6.21
Myoglobin (fragment III)	2.51

Gels were run using different reservoir buffers. The anode buffer consisted of 0.2 M Tris-HCl (pH 8.9) whilst the cathode buffer contained 0.1 M Tris-Cl (unbuffered), 0.1 M tricine, SDS (0.1% w/v). A constant 90V was passed across the gel, which changed in resistance during the run from 30 mA to 8mA.

M1.2 Staining SDS-PAGE gels.

M1.2.a Coomassie stained gels

Coomassie brilliant blue R dye was dissolved in methanol (45% v/v), acetic acid (10% v/v) to a concentration of (0.25% w/v) and filtered through Whatman No. 1 paper. Gels were soaked in dye solution with agitation for 1 hour and then transferred to destain solution [methanol (30% v/v), acetic acid (10% v/v)]. After incubation with agitation proteins became visible within 2 hours. This destaining procedure could be shortened in time by incubation at 50°C.

M1.2.b Silver stained gels

Gels were incubated with methanol (50% v/v) for at least 3 hours followed by incubation with Solution C (see below) for 15 minutes with constant gentle agitation.

Solution A: 0.8g silver nitrate, 4ml water.

Solution B: 21ml sodium hydroxide (0.36% w/v), 1.4ml 14.8 M ammonia

Solution C: Solution A added to Solution B dropwise with stirring, then made to 100ml volume with water.

Gels were then washed for 5 minutes with copious volumes of deionised water and then developed in fresh solution D (see below) for 10-15 minutes. Gels were then washed in water and fixed in destain.

Solution D: 2.5ml citric acid (1% w/v), 0.25ml formaldehyde (38% v/v), diluted to 500ml with water.

M1.3 Purification of a previously characterised Ricinus 2S albumin

The protein was purified according to the method of Li et al (1977) with modifications by McGurl (1986). All steps were performed at 4°C unless otherwise stated. Care was taken in both the handling and grinding of seeds and the subsequent purification to avoid contact with allergens and cytotoxic agents such as ricin.

30g dry castor bean seeds were dehusked and ground with liquid nitrogen in a pestle and mortar positioned within a fume hood. Powder was resuspended in 2-3 volumes 50mM imidazole-HCl (pH 7.0) and left for 30 minutes. The material was expressed through eight layers of muslin and the supernatant separated from lipid and insoluble material by centrifugation at 15,000 rpm for 10 minutes in a Sorvall 8 x 50 rotor. The supernatant proteins were precipitated by addition of ammonium sulphate to 80% saturation (516 g/l), and collected by centrifugation at 10,000 rpm for 10 minutes in a Sorvall 8 x 50 rotor, whereupon the supernatant and residual floating lipid were removed. The pellet was resuspended in 5 mls 50mM imidazole-HCl (pH 7.0) and subjected to an ether extraction, the bottom aqueous layer being retained.

On addition of sucrose to 5% (w/v) the proteins were layered onto a Sephadex G-50 column (2.5 x 45 cm) preequilibrated in 50mM imidazole (pH 7.0). The column was eluted at 25ml/hour in the same buffer and 3ml fractions collected. Absorbance at 280nm and SDS-PAGE was carried out on all fractions. The 2S albumin peak [peak 2, McGurl (1986)] was ammonium sulphate precipitated at 80% saturation, collected by centrifugation, redissolved in 5mls 10mM Tris-HCl (pH 7.0) and dialysed twice against 5L

of the same buffer. The sample was then layered onto a DEAE-Sephacel column (2.5 x 15 cm) equilibrated in and eluted in the same buffer. A linear salt gradient [0-300mM NaCl in 10mM Tris-HCl (pH7.0)] eluted the remaining proteins. Each 3ml sample was analysed for absorbance at 280nm and analysed by SDS-PAGE. Conductivity of every tenth fraction was measured to verify the linear salt gradient. The fraction containing the characterised albumin [peak 1, McGurl (1986)] was ammonium sulphate precipitated at 80% saturation, collected as before, resuspended in 10mM Tris-HCl (pH 7.0), and dialysed against same as before. The sample was layered onto a CM-52 cellulose column (2.5 x 12 cm), preequilibrated in dialysis buffer, run in and washed with one column volume of buffer at a flow rate of 0.3 ml/minute and eluted with a linear salt gradient as before. Samples were analysed as before. The major absorbance peak [McGurl (1986)] was precipitated, pelleted and resuspended as before and dialysed against 5L of distilled water twice. The volume was reduced to 2mls using Aquacide and freeze dried. The protein was weighed and resuspended in a fixed volume of water to a known concentration. All contaminated equipment was soaked overnight in water containing 2,500 ppm free chlorine to denature ricin and allergens.

M1.4 Protein body extraction and 2S albumin purification

This was performed according to the method of Tully and Beevers (1976), modified by Westby (Warwick, UK). 40g dry castor bean seeds were dehusked and ground in a mortar and pestle with 40 mls glycerol. The homogenate was expressed through eight layers of muslin and the filtrate centrifuged at 8,600 rpm for 15 minutes at 15°C in a Sorvall 8 x 50 rotor. The pellet containing intact protein bodies was resuspended in glycerol and recentrifuged. This was repeated three times. The protein body pellet was then lysed with 5 mls prechilled 25mM Na/PO_4 (pH 7.0) [39mls 0.2M NaH_2PO_4, 61mls 0.2M Na_2PO_4, 700mls H_2O] and the suspension centrifuged at 11,600 rpm to pellet membranes and crystalloid proteins. The supernatant, representing the soluble matrix proteins of the protein bodies, was then dialysed against 5L 10mM Tris-HCl (pH 7.0) overnight twice, and reduced in volume using Aquacide. The solution could then be frozen or kept at 4°C until required.

For experiments investigating enzymatic components then protein bodies were lysed in citrate-phosphate buffer (pH 4.6):

0.1M citric acid (26.7% v/v), 0.2 M Na_2HPO_4 (23.3% v/v).

The 2S albumins were efficiently separated from the 7S lectins by layering onto a Sephadex G-50 column (M1.3) and eluting with 10mM Tris-HCl (pH 7.0). The 7S lectins eluted rapidly through the matrix whilst the 2S albumins were retarded and were collected later. The 2S albumins were dialysed against 5L distilled water twice overnight and reduced in volume using Aquacide. These were then frozen at -20°C.

M1.5 Gel extraction of proteins and their purification.

Samples were run on a suitable SDS-Page gel and placed in Coomassie blue stain for 15 minutes. Destaining was performed with agitation at room temperature as described but for only 15-30 minutes maximum. The gel was illuminated from below and bands excised and placed in glycine reservoir buffer (M1.1.a). The bands were chopped into 1 cm length slices and placed into a Schleicher and Schuell Biotrap apparatus. Proteins and dye were electrophoresed into the trap using glycine reservoir buffer at 200V in accordance with the instructions from the manufacturers. The trap contents were placed in an Eppendorf tube and freeze dried overnight. The lyophilisate was resuspended in 60 ul distilled sterile water containing 0.1mg Coomassie blue and layered onto a desalting Sephadex G-15 column (14 x 0.7 cm) equilibrated in water, and eluted with water at a rate of 30mls/ hour. The blue dyefront was collected in 1ml samples and the remaining fractions collected. All samples were freeze dried overnight. Blue fractions containing no salts were resuspended, pooled and an aliquot analysed by SDS-Page. Latter fractions were checked for the presence of salts in large amounts, to indicate correct desalting.

M1.6 Western blotting

The method of Western blotting is based on the method of Towbin et al (1979). Gels, run with prestained markers (see M1.1.a) were immediately placed into transfer buffer and layered with nitrocellulose paper. Surrounded by Whatman 3MM paper and felt pads the gel/ nitrocellulose was placed in a Bio-Rad Transblot apparatus with the gel facing the cathode. The apparatus was filled with transfer buffer and the current applied. For transfer overnight a voltage of 30V was applied, for a quicker, 2 hour transfer, 60V was used with cooling.

Unless otherwise stated all incubations took place at room temperature with gentle agitation. All solutions were based on phosphate buffered saline (PBS). Once transferred, non-specific antibody binding sites were blocked by incubating the nitrocellulose filter with 150 mls Solution A for at least one hour. This was replaced by 10 mls solution A containing 50 ul rabbit antibody raised towards the characterised 2S albumin (a generous gift from McGurl) followed by incubation for three hours or overnight. The filter was then washed five times with Solution B and replaced with 10 mls Solution A containing 33 ul biotinylated protein A from the Amersham Western blotting kit, as directed. This was incubated for one hour followed by washing five times with solution B. 10mls solution C containing 33 ul streptavidin peroxidase complex from the Amersham kit were added to the filter and incubated for 30 minutes, followed by two washes with solution C and two washes in PBS. The filter was then washed in solution D and replaced with 10 mls solution D containing 6mg diaminobenzidine and 15 ul hydrogen peroxide (100

volumes). The filter was incubated for 15 minutes then washed in water, dried and photographed.

Solutions	A	B	C
Tween-20 (v/v)	0.1%	0.1%	1%
'Marvel' (w/v)	5%	-	-

Solution D and transfer buffer do not contain PBS

Solution D: 50mM Tris-HCl (pH 7.5), NaCl (0/9% w/v).

Transfer buffer: 25mM Tris-HCl (pH 8.3), 192mM glycine, methanol (20% v/v)

M2.1 Poly-A$^+$ RNA extraction from developing castor bean seeds.

All equipment was siliconised, washed thoroughly, autoclaved, baked at 300^{o}C for two hours (where possible) and handled with protective gloves to avoid contamination with ribonucleases. All solutions were made with water treated with DEPC. Similar care in the handling of the seeds was employed as described in M1.3. All handling of the preparation was carried out at 4^{o}C unless otherwise stated, and for the minimum amount of time possible.

Castor bean plants (Ricinus communis) were grown from seed in John Innes No. 1 compost in a greenhouse. They were maintained at 20^{o}C and illuminated with sodium lamps (10,000-12,000 lm m^{-2}) with a light : dark regime of 16:8 hours (Lamb, 1984). 30g of dwarf, post-testa castor bean seeds (stages D and E as described by Roberts and Lord, 1981a) were frozen in liquid nitrogen and ground in an Atomix blender within a fume hood. The powder was homogenised for two minutes in 2-2½ volumes of extraction buffer:

Extraction buffer

Tris-HCl (pH 9.0)	50mM
sodium chloride	150mM
EDTA	5mM
SDS	5% (w/v)

An equal volume of phenol/chloroform was added to the homogenate and left to stir for 5 minutes. Phases were separated by centrifugation for

5 minutes at 3,000 rpm in an MSE 6 x 300 swingout rotor. The aqueous phase was collected and the organic phase subjected to back extraction with ½ volume washing buffer (20mM Tris-HCl (pH 9.0), 2mM EDTA). The aqueous phases were pooled and the phenol/chloroform extraction repeated. The aqueous phase was made to 0.2M concentration sodium chloride and 2 volumes ice cold 95% ethanol were added, mixed and the whole left at -20°C overnight.

The precipitate was collected by centrifugation at 10,000 rpm at 2°C for 15 minutes in an MSE 6 x 300 fixed angle rotor. The pellet was washed five times in 3M sodium acetate (pH 5.5), being collected between washes by centrifugation at 10,000 rpm at 2°C for 30 minutes in a Sorvall 8 x 50 rotor. The washed pellet was then redissolved in 10 mls 0.3M sodium chloride and RNA precipitated overnight on the addition of 2 volumes 95% ethanol.

The RNA was collected by centrifugation at 8,000 rpm at 2°C for 20 minutes in a Sorvall 8 x 50 rotor and dissolved in Buffer A:

Buffer A

sodium chloride	400mM
Tris-HCl (pH 7.4)	20mM
SDS	0.2% (w/v)

The absorbance at 260 nm was recorded and the concentration of RNA adjusted so as not to exceed 4 mg/ml, where one A_{260} unit equals 40ug RNA. 60mg RNA was mixed with every 6ml (1-2g dry weight) oligo dT cellulose preswollen in buffer A. Mixing was allowed to proceed by gentle rotation on an angled revolving turntable. The oligo dT cellulose

was then collected by centrifugation at 3,000 rpm for 1 min at room temperature in an MSE benchtop centrifuge and washed three times with buffer A. After washing three times with buffer B:

Buffer B

sodium chloride	200mM
Tris-HCl (pH 7.4)	20mM
SDS	0.1% (w/v)

the matrix was transferred to a column (5 x 0.8cm) connected to an ISCO spectrophotometer and chart recorder. Buffer B was passed down the column until a steady baseline was achieved then poly A^+ RNA was eluted with 20mM Tris-HCl (pH 7.4) preheated to 50°C, and collected. This was made to 0.2M concentration sodium chloride and precipitated overnight with 2 volumes 95% ethanol. On centrifugation as before the pellet was washed twice with 70% ethanol, vacuum dried, and resuspended in sterile distilled water to a concentration of 1 mg/ml. Poly A^+ RNA was stored at -80°C.

Poly A^+ RNA was also extracted from germinating castor bean seeds under exactly the same conditions. Germinating castor bean seeds were prepared by imbibing overnight in running water, then growing at 30°C for three days in darkness.

M2.2 In vitro translations

M2.2.a Rabbit reticulocyte lysate translation

The method of Pelham and Jackson (1976) was employed with mixes kindly donated by M. May (Warwick, UK). Translations occurred in reticulocyte lysates that had been previously nuclease treated. Reactions were in 20 ul volumes and were incubated for one hour at 30°C. When required dog pancreatic microsomal membranes were included according to the manufacturer's instructions.

Translation:	ul
Reaction mix	7
L-[^{35}S] methionine	2
Poly A$^+$ RNA	1
Lysate	10

Methionine was labelled to a specific activity >30TBq/mmol.

Reaction mix	ul
2M potassium chloride	137.5
40mM magnesium acetate	117.5
10mM Tris-HCl (pH 7.4)	80
Amino acid mixture	125
Energy mixture	155
H_2O	492.5

Amino acid mixture contained all L-amino acids except methionine at 2mM concentration at pH 8.3.

Energy mixture

ATP	8mM
GTP	1.6mM
creatine phosphate	80 mg/ml
Tris-HCl (pH 7.5)	0.2M

M2.2.b Wheat germ translation

This was performed according to the method of Marcu and Dudock (1974) and both the wheatgerm extract and reaction mix were generous gifts of M. Westby (Warwick, UK). Translations were performed in 25ul reactions and incubated at 27°C for one hour. Prior to use wheatgerm lysate was spun for 5 minutes at 4°C in an MSE microfuge at 13,000 rpm, and only the supernatant used in the reaction.

Translation	ul
Reaction mix	4.8
L-[^{35}S] methionine	2
H_2O	9.7
Poly A^+ RNA	1
Lysate supernatant	7.5

Methionine was labelled to a specific activity >30 TBq/mmol.

Reaction mix	ul
1M Hepes-KOH (pH 7.6)	140
0.1M ATP	100
0.4M creatine phosphate	200
creatine phosphokinase (10 mg/ml)	40
spermine (1.5 mg/ml)	200
0.1M dithiothreitol	170
2mM GTP	100
2mM amino acids (no methionine)	125
0.1M magnesium acetate	40
1.0M potassium acetate	840

Where L-[^{35}S] cysteine was used to label translational products, L-[^{35}S] methionine was excluded and the amino acid pool used contained all 2mM amino acids except cysteine.

M2.2.c Estimation of translation efficiency

Translations were assessed by their ability to incorporate radiolabel into TCA precipitable material which could be collected and counted by scintillation. 2ul of translation products were placed onto Whatman 3MM paper and dried. The paper was then washed for 10 minutes in ice cold TCA (10% w/v). The paper was then transferred to boiling TCA (5% w/v) for 10 minutes and then washed twice for two minutes in TCA (5% w/v) at room temperature. The filter was then washed in ethanol for two minutes, dried and placed in non-aqueous scintillant and counted.

M2.2.d Immunoprecipitation of translation products

Translation products were diluted to 100 ul with water and solubilised by the addition of 100 ul buffer A. After spinning for 10 minutes the supernatant was removed and added to 2 ul null serum followed by incubation for 45 minutes. 30 ul swollen protein A Sepharose was added and incubated for 45 minutes with rotation. The matrix was precipitated by centrifugation and to the supernatant was added 2 ul rabbit serum containing antibodies raised towards the previously characterised Ricinus 2S albumin. This was incubated for 45 minutes followed by the addition of 30 ul swollen protein A Sepharose and incubation for 45 minutes with rotation. Matrix was recovered by brief centrifugation and washed 3 times in buffer B, two times in buffer C and once in buffer D. The matrix was then boiled in sample buffer and loaded onto the appropriate gel. All incubations were performed at room temperature and all centrifugations performed in an MSE microfuge at 13,000 rpm at room temperature.

Solutions	A	B	C	D
sodium chloride	150mM	150mM	500mM	-
EDTA	2mM	2mM	2mM	-
Tris-HCl (pH 7.4)	20mM	20mM	20mM	10mM
Nonidet P-40 (v/v)	1%	0.2%	0.2%	-

M2.2.e Fluorography

All SDS-Page gels containing [^{35}S] labelled products were subjected to fluorography by immersion in Amplify solution for 20 minutes prior to drying in accordance with the manufacturers instructions.

M2.3 RNA electrophoresis

M2.3.a Formaldehyde agarose gel

For a 1.5% agarose gel, 1.5g agarose was boiled with 75 ml water and 10ml 10x MOPS buffer. Once the agarose was melted and the solution cooled to 60°C 15ml formaldehyde (filtered through Whatman No.1 paper) was added and the solution poured into a mould with well-former to set, within the fume cupboard.

10x MOPS buffer:

MOPS	0.2M
sodium acetate	0.05M
EDTA	0.01M

adjusted to pH 7.0 with sodium hydroxide and autoclaved.

RNA was adjusted to 5 ul volume with water to which was added 15 ul denaturing solution:

Denaturing solution	ul
Deionised formamide	500
10x MOPS buffer	100
Formaldehyde	150

RNA and denaturing solution were heated to 60°C for 5 minutes and then 2 ul loading mix (50% glycerol, 0.2% bromophenol blue) was added. The sample was immediately loaded onto a formaldehyde agarose gel immersed in 1x MOPS buffer within the fume cupboard and the current applied. Gels typically took 4 hours when run at 100mA.

Once the gel was run it was soaked in glycine (10% w/v) for 30 minutes, then ethidium bromide (1 mg/l) for 30 minutes, then photographed under ultraviolet light using Polaroid 667 film.

M2.3.b Formamide agarose gel

For a 1.5% agarose gel, 1.5g agarose was boiled and melted in 40mls water. Upon cooling to 60°C 50mls formamide and 10 mls 10X TEP buffer were added and the solution poured into a mould with well-former to set.

10x TEP buffer	mM
Tris-HCl (pH 8.0)	36
NaH_2PO_4	30
EDTA	2

Gels were immersed in but not covered by 1X TEP buffer. Samples of RNA in 3 ul volume were added to 20 ul formamide (60% v/v), 1x TEP buffer. The samples were heated to 65°C for 5 minutes then quenched in ice. 3 ul loading buffer was added as before and samples applied to the gel. Gels were run at 20 mA for 2½ hours until the bromophenol blue dyefront had reached 2/3 down the gel. The gel was then stained in ethidium bromide and photographed. Gels were run with 6 ug RNA size markers from BRL.

M2.4 Northern blotting and probing of RNA blots

Gels from either the formaldehyde or formamide method were soaked twice in 20x SSPE buffer for 30 minutes.

20x SSPE Buffer

sodium chloride	3.6M
sodium phosphate (pH 7.7)	0.2M
EDTA	2 mM

Gels were then blotted as described in the Amersham 'Membrane Transfer and Detection Methods' manual following the method of Thomas (1980). Blots were prehybridised, hybridised and washed as also described therein. Following washing blots were exposed to X-ray film by autoradiography (M3.4) RNA size markers were stained on filters by immersion in methylene blue as described by Maniatis et al (1982).

M3 DNA TECHNIQUES

M3.1 General DNA techniques

Unless otherwise stated all centrifugation was performed in Eppendorf tubes within an MSE microfuge at room temperature and at 13,000 rpm.

M3.1.a DNA precipitation

DNA was precipitated from aqueous solution on the addition of both salt and alcohol. In particular the salt of routine choice was sodium acetate diluted from 3M stock to 0.3M in the solution containing the DNA. Other salts used however included sodium chloride, from a 2M stock to 0.2M working concentration, and ammonium acetate from a 7.5M stock to 2.0M. Alcohol added was usually two volumes of 95% ethanol prechilled to -20°C, but this was sometimes exchanged for 1 volume isopropanol. Upon mixing the solution, the DNA was allowed to precipitate by incubation at -80°C for 15 minutes or -20°C for 1 hour. After centrifugation for 10 minutes, the supernatant was removed and replaced with prechilled 70% ethanol, followed by another centrifugation step for only one minute. This 70% ethanol wash was repeated. Finally the pellet was dried in a Speedivac centrifuge and resuspended in the buffer of choice.

M3.1.b Phenol/chloroform extraction

Proteins were removed from solutions containing DNA prior to ethanol precipitation by vortexing with an equal volume of phenol/chloroform. The DNA solution was made to at least 50 ul in volume with water prior to the extraction. After vortexing, the phases were separated by centrifugation for 3 minutes. The upper aqueous layer was removed and

vortexed with an equal volume of chloroform, and centrifuged for 3 minutes. The upper aqueous layer was removed and vortexed with an equal volume of ether. The upper ether phase was removed, and the lower aqueous phase then ethanol precipitated as described.

M3.1.c Minipreparation of plasmid DNA

DNA was prepared from cultures of Escherichia coli transformed with the plasmid by the quick boiling method of Holmes and Quigley (1981) adapted by R. Spooner (Warwick, UK). Bacteria grown up overnight in 10 ml of L-Broth:

L-Broth	g/l
sodium chloride	10
Bacto-tryptone	10
yeast extract	5

pH to 7.3 with 1M sodium hydroxide.

containing ampicillin at a concentration of 0.1 mg/ml were pelleted by centrifugation at 7,000 rpm for 1 minute in a Sorvall 8 x 50 rotor. The pellet was resuspended in 180 ul of SET buffer:

SET buffer	
sucrose	20% (w/v)
Tris-HCl (pH 8.0)	0.1M
EDTA	50mM

and transferred to Eppendorf tubes. 180 ul of SET containing lysozyme (4 mg/ml) was added and left at room temperature for 5 minutes after mixing. 300 ul Triton X-100 was added, mixed and the sample boiled for

two minutes followed by quenching on ice. The lysate was then centrifuged for 30 minutes and the supernatant placed in a fresh tube. 300 ul 7.5M ammonium acetate was added and left for 20 minutes at 4°C after mixing. The sample was centrifuged for ten minutes and to the supernatant was added 630 ul isopropanol to precipitate the nucleic acid. The precipitation was performed as usual. DNA was resuspended in 50 ul water and 5 ul used in restriction analysis (M3.1.f).

M3.1.d Large scale preparation of plasmid DNA

Using the same method for the minipreparation of plasmid DNA, 400 ml of an overnight culture was centrifuged in a Sorvall 6 x 250 rotor at 5,000 rpm for 5 minutes. The cell pellet was resuspended in 14.4 ml SET buffer containing 30 mg lysozyme and left for 5 minutes at room temperature. 12 ml Triton X-100 was added and the solution boiled over an open flame in a 1L conical flask, followed by boiling for 1 minute in a water bath. The lysed cells were quenched on ice for 1 minute then centrifuged at 21,000 rpm at 4°C in a Sorvall 8 x 50 rotor for 20 minutes. To the supernatant was added ¼ volume of 7.5M ammonium acetate. This was left for 20 minutes at 0°C. The solution was centrifuged at 15,000 rpm at 4°C in a Sorvall 8 x 50 rotor for 10 minutes and to the supernatant was added 0.7 volume isopropanol. This was left at -80°C for 10 minutes then nucleic acid recovered by centrifugation as before. The pellet was resuspended in 4 ml TE buffer (10mM Tris-HCl (pH 7.0), 1mM EDTA). To this was added 0.5 ml ethidium bromide (5 mg/ml) and 4.3g caesium chloride. The solution was loaded into a Beckman opaque Quickseal tube and heat sealed. The column was centrifuged at 52,000 rpm at 20°C in a Vti65 rotor for 16 hours. Covalently closed circular plasmid DNA was

extracted from the gradient by hypodermic syringe and subject to extraction four times with isopropanol saturated with caesium chloride and TE buffer to remove ethidium bromide. The remaining solution was diluted fivefold with TE buffer and ethanol precipitated as usual. DNA was diluted in the appropriate volume of TE and the absorbance at 260nm was measured. The DNA was then diluted to known concentration, where 1 A_{260} unit equals 50 ug DNA.

M3.1.e Electrophoresis of DNA using agarose

Two types of agarose gel electrophoresis were employed depending on the required study: neutral and alkaline.

Neutral gels

The required weight of agarose (0.8-1.5g/100ml typically) was melted by boiling in 90 ml water. To this was added 10ml 10x TAE or TBE buffer:

10x TBE buffer	g/L	10x TAE buffer	g/L
Tris base	108	Tris base	48.5
EDTA	9.5	EDTA	3.7
boric acid	55	sodium acetate	4.1
pH 8.4 with boric acid		pH 7.9 with acetic acid	

and 10 ul ethidium bromide (10 mg/ml). The molten agarose was then poured into a gel former with well forming comb and allowed to set at room temperature. The gel was placed in either 1x TAE or TBE buffer respectively and the comb removed. DNA samples were mixed with loading

buffer, loaded into the wells and a current applied. TBE buffered gels were run at 100V constant whilst TAE buffered gels were run at 50V.

Loading buffer (5X concentration)	
ficoll 4000	15% (w/v)
EDTA	1mM
bromophenol blue	0.1% (w/v)

DNA size markers were routinely run with agarose gels, the most common being lambda DNA restriction endonuclease treated with Hind III and Eco RI. or just with Hind III (M3.1.f). Before loading, markers were heated to 65^{o}C for 5 minutes to melt lambda cos ends.

Hind III	Hind III/Eco RI
23.130	21.226
9.416	5.418
6.557	4.973
4.361	4.277
2.322	2.027
2.027	1.904
0.564	1.584
0.125	1.330
	0.983
	0.831
	0.564
	0.125

Gels were photographed under ultraviolet light using Polaroid 667 positive film.

Alkaline agarose gels

These gels were run when single stranded DNA was under investigation, typically when the products of cDNA synthesis reactions were being examined. 1g agarose was melted in 90ml water as before and cooled to 55^{o}C when 10ml 10x buffer was added (300mM sodium hydroxide, 20mM EDTA). The gel was poured without ethidium bromide, and when set immersed in 1x alkaline buffer. Samples were loaded with an equal volume sample buffer:

Sample buffer (2X concentration)

Tris-acetate (pH 8.0)	80mM
EDTA	8mM
Glycerol	40% (v/v)
Bromocresol green	0.02% (w/v)

Gels were run with circulating buffer at 30V initially then 50-60V until the dye was 2/3 down the gel. Gels were then fixed for 30 minutes in 0.1M ammonium acetate, then stained in ethidium bromide (0.5 ug/ml) for 30 minutes for visualisation of bands and photography. Gels were then dried onto cling film at 73^{o}C under vacuum and exposed to X-ray film by autoradiography.

M3.1.f Restriction endonuclease digestion of DNA

DNA was subject to restriction enzyme digestion in 20 ul reaction volumes for every 1-2ug DNA. DNA was diluted to 19 ul in 1X buffer appropriate to the enzyme (buffer being supplied with the enzyme by the manufacturers). 1 ul enzyme (5-10 U generally, where 1 U digests 1 ug DNA in 1 hour completely) stored at -20^{o}C was then added and incubated

at 37°C (30°C for Sma I) for at least 1 hour. If more than one enzyme was being used simultaneously, then the total enzyme concentration was never allowed to exceed 10% of the total volume of the reaction. Samples were either then frozen for later use, prepared for gel electrophoresis, or heat inactivated by incubation at 65°C for 10 minutes if the DNA was required for ligations. The following enzymes were routinely used:

Enzyme	Reference
BamH I	Wilson and Young (1975)
Eco RI	Yoshimori et al (1974)
Hind III	Old et al (1975)
Pvu II	Gingeras et al (1981)
Sac I	Roberts (1983)
Sma I	Endow and Roberts (1977)
Sph I	Fuchs et al (1980)

M3.1.g Transformation of Escherichia coli

The method of Hanahan (1985) was employed to make Escherichia coli cells competent and to transform them with DNA. 2 strains of E.coli were used most: JM101 [Δlac, pro, thi, supE, F', traD36, proAB, lacI^{q}ZΔM15) was used for M13 transfection whilst DH5a (F$^-$, endA1, hsdR17 (r_k^- m_k^+), supE44, thi-1, λ$^-$, recA1, gyrA96, relA1, Ø80dlacZΔM15) was used for routine plasmid transformation.

A bacterial culture grown in 10 ml SOB medium overnight with shaking at 37°C was used to inoculate (dilution to 1%) a fresh prewarmed flask of SOB medium. Cells were grown till they reached an optical density of 0.45-0.48 at 660nm. These were cooled on ice for 10 minutes and

centrifuged gently at 2,500 rpm for 5 minutes at 4°C in a Sorvall 8 x 50 rotor. The cells were then resuspended in 1/3 volume prechilled RF1 buffer, and left on ice for 15 minutes. The cells were pelleted as before, resuspended in 1/12 volume prechilled RF2 buffer and left on ice for 15 minutes. The cells, ready for transformation, were aliquotted into 200 ul volumes and stored at -80°C after rapid freezing in liquid nitrogen. When required cells were allowed to thaw on ice, DNA was added and allowed to incubate for 10-60 minutes. Cells were then heat-shocked for 90 seconds at 42°C and chilled on ice for 2 minutes. 800 ul SOC medium was then added to the 200 ul aliquot and left to incubate at 37°C for 30 minutes. Cells were then spread onto the appropriate solid media and incubated overnight at 37°C.

SOB medium

Bacto-tryptone	2% (w/v)
yeast extract	0.5% (w/v)
sodium chloride	10mM
potassium chloride	2.5mM
magnesium chloride	10mM
magnesium sulphate	10mM

pH 6.8-7.0

SOC medium

Same as SOB medium but contained 20mM glucose.

RF1 buffer

rubidium chloride	100mM
manganous chloride	50mM
potassium acetate	30mM
calcium chloride	10mM
glycerol	15% (v/v)

Final pH 5.8 (acetic acid), filter sterilised.

RF2 buffer

MOPS	10mM
rubidium chloride	10mM
calcium chloride	75mM
glycerol	15% (v/v)

Final pH 6.8 (sodium hydroxide), filter sterilised.

M3.1.h Phosphatasing vector DNA

1ug restriction endonuclease digested DNA was resuspended in 17 ul water to which was added 2 ul 10x CIP buffer:

CIP buffer (10x)

Tris-HCl (pH 10.5)	0.5M
Spermidine	10mM
EDTA	1mM

1 ul of calf intestinal alkaline phosphatase (1 U/ul, where one unit catalyses the hydrolysis of one umol p-nitrophenyl phosphate per minute at 37^{o}C) was added and incubated at 37^{o}C for 30 minutes followed by the further addition of 1 ul of enzyme and incubation at 37^{o}C for 30 minutes. Enzyme was inactivated by incubation with 2 ul 0.1M

trinitriloacetic acid at 70°C. This method was applied to DNA where 5' overhangs were generated by the restriction enzyme digest. In the case of blunt ends and 3' overhangs, both 30 minute incubations were divided into 15 minutes at 37°C and 15 minutes at 56°C.

M3.1.1 Ligation of DNA

Ligations were carried out as follows (volumes in ul):

10X Ligation buffer	2
100mM dithiothreitol	2
10mM rATP	2
Vector DNA	2
Fragment DNA	2
H_2O	9
T4 DNA ligase (5U/ul)	1
Ligation buffer (10x)	
Tris-HCl (pH 7.4)	0.5M
magnesium chloride	0.1M
spermidine	10mM
BSA	1 mg/ml

Where blunt ended DNA was being ligated, 2 ul 10mM hexamine cobalt chloride was included instead of 2 ul water as according to Rusche and Howard-Flanders (1985). One unit of T4 DNA ligase is defined according to the manufacturers as the amount of enzyme which joins more than 90% of 6 ug Hind III cleaved lambda DNA fragments in 30 minutes at 16°C. Reactions were left overnight at 16°C or for 3 hours at room

temperature. Reactions were then used to directly transform competent E.coli as described (M3.1.g).

M3.1.j DNA isolation from agarose gels

DNA was isolated by using low melting point agarose gels run in TAE buffer (M3.1.e) following the method of Wieslander (1979). DNA bands visualised by ultraviolet light were excised and melted at 65°C for 10 minutes in 150 ul TE buffer (3.1.d). Phenol saturated with TE buffer was heated to 65°C and 200 ul added to the molten agarose. After thorough vortexing the phases were separated by centrifugation for 5 minutes. The upper aqueous phase was subject to another phenol extraction, 10 ug E.coli tRNA was added and then ethanol precipitated as usual.

M3.1.k Phage Lambda DNA minipreparation

Bacteriophage lambda was handled exactly as described in Maniatis et al (1982) for techniques such as preparation of plating bacteria, plating lambda, picking plaques, and preparing stocks of lambda from single plaques by plate lysate. A different method was used for the minipreparation of lambda DNA. This was described at the EMBO course on Plant Molecular Biology held in Gent, Belgium in August 1987 and kindly passed on by J. Tregear (Warwick, UK).

Plaques were picked with a Pasteur pipette and resuspended in phage storage medium:

Phage storage medium

Tris-HCl (pH 7.2)	10mM
magnesium sulphate	10mM
Bacto-gelatin	0.02% (w/v)

for 30-60 minutes at room temperature. 50 ul of the supernatant was mixed with 50 ul E.coli growing at late logarithmic phase and 50 ul 10mM magnesium chloride, 10mM calcium chloride. Lambda was allowed to absorb for 10 minutes then 4ml of NZ medium was added and shaken vigorously at 37°C for 6 hours or until bacterial lysis was apparent.

NZ medium	g/L
NZ amine	10
sodium chloride	5
magnesium chloride ($6H_2O$)	2

2 ml of the lysate was transferred to a Falcon tube containing 0.4 ml Buffer A and incubated at 70°C for 30 minutes.

Buffer A

Tris-HCl (pH 9.0)	0.5M
EDTA	0.25M
SDS	2.5% (w/v)

0.8 ml 5M potassium acetate was added and incubated on ice for 15 minutes. The sample was centrifuged at 10,000 rpm for 20 minutes at room temperature using a Sorvall 8 x 50 rotor. The supernatant was poured through sterile Miracloth and the nucleic acid precipitated with

ethanol. The nucleic acid pellet was resuspended in 0.4 ml 2M ammonium acetate and precipitated with ethanol as usual.

M3.1.1 SP6/T7 polymerase in vitro transcription

DNA was subcloned into pGem3blue and a large scale preparation made of the construct (M3.1.d). 10 ug was linearised by using a restriction enzyme contained within the polylinker that was not contained within the subcloned DNA and that was distal to the polymerase to be used in the transcription. Usually DNA was subcloned in an orientation so that transcription with T7 RNA polymerase produced a transcript that would contain the desired open reading frame for subsequent translation. After restriction enzyme linearisation the DNA was phenol/chloroform extracted, ethanol precipitated and resuspended in 5 ul water.

Transcription reaction	ul
Premix	12
RNAsin (25u/ul)	0.5
H_2O	2.5
5' cap [m7G(5')ppp(5')G] 5mM	1
DNA (2 ug/ul)	2
RNA polymerase (20U/ul)	2

Premix	ul
BSA (10 mg/ml)	3
100mM dithiothreitol	6
10mM ATP, CTP, UTP (each)	3
1mM GTP	3
transcription buffer (10x)	6
H_2O	20.4
10x transcription buffer	mM
spermidine	20
HEPES-KOH (pH 7.5)	400
magnesium acetate	60

The reaction was incubated at 40°C for 30 minutes, followed by the addition of 1 ul 8mM GTP in 20mM HEPES-KOH (pH 7.5) and a further incubation period of 30 minutes at 40°C. One unit of RNA polymerase catalyses the incorporation of 1 nmol of nucleoside triphosphate into acid precipitable products in 60 minutes at 37°C. 1 unit RNAsin (human placental ribonuclease inhibitor) is that which inhibits by 50% the activity of 5ng ribonuclease A. Transcripts were used directly in in vitro translation systems using 1 ul reaction products instead of poly A^+ RNA (M2.2).

M3.2 cDNA synthesis

cDNA was produced using the S1 nuclease method detailed by a number of workers: Verma *et al* (1972), Efstratiadis *et al* (1976), Buell *et al* (1978) and Wickens *et al* (1978). All reactions were performed according to the optima for the enzymes being used as directed by the manufacturers. cDNA was made from both developing and germinating castor bean seed poly A^+ RNA.

M3.2.a First strand synthesis

Reaction mixture	ul
RT mixture	60
Poly A^+ RNA (1 ug/ul)	10
AMV reverse transcriptase (25U)	2
[a-^{32}P]dCTP	2
H_2O	26

This was incubated at 42°C for 45 minutes after which was added:

Dilution mix (2x)	50
AMV reverse transcriptase (25U)	2
H_2O	48

This was incubated at 45°C for 45 minutes. 1 ul was taken for subsequent gel analysis and 0.5 ul taken for incorporation analysis.

RT mixture	ul
1M Tris-HCl (pH 8.3)	20
1M magnesium chloride	4
2M potassium chloride	20
10mM dATP, dGTP, dTTP (each)	40
10mM dCTP	10
Oligo dT_{12-18} (1 ug/ul)	24
1M dithiothreitol	4
H_2O	38
(enough for 4 reactions).	

Dilution mixture (2X)	ul
1M Tris-HCl (pH 8.3)	2
1M dithiothreitol	2
10mM dCTP	10
H_2O	186

[α-^{32}P]dCTP was of specific activity 110TBq/mmol. One unit of reverse transcriptase catalyses the incorporation of 1.0 nmol of [^{3}H]-dTMP into acid precipitable product in 10 minutes at 37^{o}C.

M3.2.b Second strand synthesis

The first strand reaction products were boiled for 3 minutes and quenched rapidly in ice water and centrifuged briefly. The supernatant was transferred to a fresh chilled tube to which was added 39 ul DNPase mixture, 1 ul label as before and 24 units E.coli DNA polymerase I (nuclease free preparation). This was incubated at 20^{o}C for six hours. 1.5 ul was removed for subsequent gel analysis, and 3 ul for

incorporation analysis. The reaction was terminated by the addition of 20 ul 100mM EDTA.

DNPase mixture	ul
10mM dATP, dGTP, dTTP (each)	10
10mM dCTP	8
1M HEPES-KOH (pH 6.9)	100
2M potassium chloride	20

1 unit of DNA polymerase 1 incorporates in 30 minutes under assay conditions 10 nmol of total nucleotides into an acid precipitable fraction.

The sample was layered onto a Biogel P-60 column (Pasteur pipette) preswollen in NET buffer. 120 ul volumes of NET were added, collected and counted in scintillation vials measuring in the ^{3}H channel. Radioactive peaks were pooled, excluding the impeded unbound label, and ethanol precipitated.

NET buffer	mM
Tris-HCl (pH 7.6)	10
sodium chloride	20
EDTA	1

M3.2.c S1 nuclease treatment

The DNA pellet was resuspended in 10ul water and subject to the following reaction (volumes in ul):

DNA	10
S1 buffer (10x)	4.5
H_2O	30.5
S1 nuclease (1u/ul)	5

S1 buffer (10x)

sodium acetate (pH 4.6)	300mM
sodium chloride	2.5M
zinc chloride	10mM
glycerol	5% (w/v)

and incubated for 4 minutes at 37°C then 15 minutes at 15°C. The reaction was terminated on the addition of 1 ul 100mM EDTA and 49 ul 10mM Tris-HCl (pH 7.6). This was then phenol/chloroform extracted and ethanol precipitated. One unit of enzyme causes 1 ug of nucleic acid to become perchloric acid soluble per minute at pH 4.6 at 37°C. The DNA was resuspended in 5 ul and 0.5 ul taken for subsequent gel analysis. An alkaline agarose gel (M3.1.e) was run of the products of first and second strand synthesis and the S1 nuclease digestion together with radioactive markers generated by the end-fill reaction (M3.3.d).

M3.2.d Estimation of counts incorporated during cDNA synthesis

Samples taken were placed in 200 ul TE buffer. 20 ul was spotted onto a Whatman GF1 filter and dried. To the remaining 180 ul was added 20 ul

BSA (10 mg/ml], followed by 50 ul TCA (50% w/v) and left on ice for 30 minutes. This was filtered onto a GF1 filter which subsequently had 20 ml TCA (10% w/v) passed through under suction. The filter was dried and both filters counted in non-aqueous scintillant. From this the incorporation of radiolabelled nucleotide could be expressed as a percentage of the total radiolabel. The mass of cDNA synthesised could then be derived (McGurl, 1986).

M3.2.e DNA polymerase fill-in reaction

S1 nuclease treated cDNA was subject to the repair mechanism of T4 DNA polymerase in order to ensure that cDNAs were blunt ended in readiness for the ligation of synthetic linkers.

Reaction mixture	ul
cDNA (100 ng/ul)	10
T4 DNA polymerase buffer (10x)	2
BSA (1 mg/ml)	2
2-mercaptoethanol (1% v/v)	1
H_2O	2.5
dATP, dCTP, dGTP, dTTP 2mM	2
T4 DNA polymerase (10U/ul)	0.5
T4 DNA polymerase buffer (10x)	
Tris-HCl (pH 8.8)	670mM
magnesium chloride	67mM
ammonium sulphate	166mM
EDTA	67 uM

The reaction proceeded at 37°C for 10 minutes then at 15°C for 30 minutes, followed by phenol/chloroform extraction and ethanol precipitation. One unit enzyme catalyses the incorporation of 10nmol total nucleotide into acid insoluble product in 30 minutes at 37°C. The DNA was resuspended in 5 ul TE buffer.

M3.2.f Eco RI methylase protection of cDNA

Reaction mixture	ul
cDNA (1 ug/ul)	1
Eco RI methylase buffer	20
0.1mM S-adenosylmethionine	2
Eco RI methylase (1 U/ul)	1

The reaction was incubated at 37°C for 15 minutes then 70°C for 10 minutes. DNA was precipitated and resuspended in 10 ul water.

Eco RI methylase buffer	mM
Tris-HCl (pH 7.5)	50
EDTA	1
dithiothreitol	5

1 unit of enzyme protects >90% of 1 ug lambda DNA in 30 minutes against cleavage by Eco RI under assay conditions.

M3.2.g Addition of synthetic linkers to cDNA

Eco RI linkers of sequence 5'd(GGAATTCC)3' were labelled by T4 polynucleotide kinase with [gamma-^{32}P] ATP (M3.3.a). These were then ligated onto cDNAs by T4 DNA ligase (M3.1.i).

Ligation reaction	ul
cDNA (200 ng/ul)	5
Kinased linkers (100 ng/ul)	10
100mM dithiothreitol	2.5
10mM rATP	2.5
10x ligase buffer	2.5
T4 DNA ligase (5 U/ul)	2

The ligation was performed at 15^{o}C overnight and terminated by heat inactivation at 65^{o}C for 5 minutes. The DNA was precipitated as before and subject to Eco RI digestion (M3.1.f). Excess linkers and unincorporated label were removed by passing the sample through a P-60 column as before (M3.2.b). cDNA was then ligated with phosphatased vector pUC19x, transformed into competent E.coli DH5a and plated onto selective medium.

M3.2.h Selection of recombinant plasmids

Bacteria transformed with pUC19x were selected by plating the direct transformation mixture on medium containing 100 ug/ml ampicillin. Bacterial colonies containing recombinant vector were distinguished from religated vector by the use of colour selection (Davies et al, 1980). The direct plating medium was supplemented with 12 ug/ml isopropyl-B-D-

thiogalactopyranoside and 50 ug/ml 5-bromo-4-chloro-3-indolyl-B-D-galactopyranoside. Bacteria transformed with recombinant plasmids were unable to utilise the substrate and their colonies thus remained white in colour compared with those transformed with religated plasmid which became blue.

M3.3 Preparation of radioactive probes

M3.3.a Kinase reaction

The technique was used for the labelling of mixed synthetic oligonucleotides for the screening of the cDNA plasmid library (Conner et al. 1983)

Reaction mixture	ul
5'-[Gamma-^{32}P] ATP	3
Oligonucleotide (10 pmol/ul)	1.5
100mM dithiothreitol	1
Kinase buffer (10x)	3
T4 polynucleotide kinase (5 U/ul)	1
H_2O	20.5

Kinase buffer (10x)	
Tris-HCl (pH 7.8)	500mM
magnesium chloride	100mM
spermidine	10mM
BSA	2 mg/ml

The reaction proceeded at 37°C for 30 minutes, was diluted in 3ml 6xSSC (M3.4) and filtered through a 0.22 um Multex filter before addition to the hybridisation solution (M3.4). One unit of enzyme catalyses the production of 1nmol of acid insoluble ^{32}P in 30 minutes at 37°C. 5'-[Gamma-^{32}P] ATP possessed a specific activity of >185 TBq/mmol.

M3.3.b Nick translation

This was used to label cloned DNA (Maniatis et al. 1975) prior to the use of oligo-labelling as the standard technique.

Reaction mixture	ul
DNA (100 ng/ul)	1
NT buffer (10x)	3
0.1mM dATP, dCTP, dTTP (each)	1
[α-^{32}P]dGTP	2
dGTP 1uM	1
DNase (50 ng/ul)	1
H_2O	21

The reaction proceeded at room temperature for 15 minutes then 1 ul E.coli DNA polymerase 1 (5U/ul) was added and incubated at 15°C for 3 hours. The reaction was passed down a P-60 column (M3.2.b), the labelled DNA being ethanol precipitated and resuspended in 100ul TE buffer before incubation at 95°C for 5 minutes and addition to the hybridisation medium.

NT buffer (10x)	mM
Tris-HCl (pH 7.8)	500
magnesium chloride	50
dithiothreitol	100

M3.3.c Oligo-labelling

This was followed according to the method of Feinberg and Vogelstein (1983). DNA to be labelled was boiled for 5 minutes in a water bath and transferred to 37°C for 5 minutes. This was then used in the reaction mixture.

Reaction mixture	ul
Oligo-labelling buffer	3
BSA (10 mg/ml)	0.6
[a-^{32}P]dGTP	1.5
DNA polymerase (Klenow) 5U/ul	0.6
DNA (10 ng/ul)	1.5
H_2O	7.8

Oligo-labelling buffer

This is made fresh from the following solutions A, B, C in the ratio 2:5:3 respectively.

Solution A	ul
2M Tris-HCl (pH 8.0)	625
5M magnesium chloride	25
H_2O	350
2-mercaptoethanol	18
0.1M dATP, dCTP, dTTP (each)	5

Solution B: 2M HEPES-NaOH (pH 6.6)

Solution C: Hexadeoxyribonucleotides (90 A_{260} units/ml) in 3M Tris-HCl (pH 7.0), 0.2mM EDTA.

The reaction was allowed to proceed at room temperature for 3-4 hours and terminated with 85 ul stop solution.

Stop solution

sodium chloride	20mM
Tris-HCl (pH 7.5)	20mM
EDTA	2mM
SDS	0.25% (w/v)

2ul was removed for incorporation analysis before the probe was boiled for 5 minutes and added to the hybridisation solution.

Incorporation analysis

1 ul probe was placed onto each of two pieces of DE-81 paper and dried. One piece was washed 5 times in 0.5M Na_2HPO_4 (5 minutes/wash) then twice in water (1 minute/wash) then twice in 95% ethanol (1 minute/wash). The filter was dried and both washed and unwashed filters counted in non-aqueous scintillant.

[a-^{32}P]dGTP had a specific activity of 110 TBq/mmol. The Klenow fragment of E.coli DNA polymerase 1 was used which possessed an activity where one unit catalysed the incorporation of 10nmol total nucleotide into acid insoluble product in 30 minutes at 37°C.

M3.3.d End-filling

This was used to label restriction fragments of DNA for use as gel markers by filling in the overhangs, created by digestion, with nucleotides containing radiolabelled nucleotide and the Klenow fragment of E.coli DNA polymerase 1.

Reaction mixture	ul
DNA (50 ng/ul)	20
NT buffer (10x) (M3.3.b)	3
2mM dATP, dCTP, dTTP (each)	1
[a-^{32}P]dGTP	1
H_2O	2
DNA polymerase 1 (Klenow) 5 U/ul	1

The reaction proceeded at room temperature for 30 minutes, was phenol/chloroform extracted and ethanol precipitated. Typically 1-2% was used for gel markers. Radiolabel activity was as before.

M3.4 Screening DNA by use of radiolabelled probe hybridisation

All methods were used exactly as found in the Amersham manual 'Membrane Transfer and Detection Methods'.

M3.4.a Preparation of DNA onto nitrocellulose filters

Colonies and plaques were blotted onto nitrocellulose filters, denatured for 7 minutes on pads soaked in denaturing solution (1.5M sodium chloride, 0.5M sodium hydroxide) and neutralised likewise in neutralising solution twice for 3 minutes.

Neutralising solution	
sodium chloride	1.5M
Tris-HCl (pH 7.2)	0.5M
EDTA	1mM

Filters were then washed in 2xSSC buffer:

SSC (20x)	M
sodium citrate	0.3
sodium chloride	3

Filters were then air dried and baked in vacuo for 2 hours.

DNA run on agarose gels was incubated in denaturing solution for 15 minutes twice and then for 30 minutes. Denaturing solution was replaced with neutralising solution for 1 hour with one change of buffer. The gel was placed upon a filter paper wick, covered with nitrocellulose, upon which was placed absorbency towels and a 1kg weight, as detailed

originally by Southern (1975). Transfer was achieved overnight using 20x SSC as transfer buffer. Filters were then air dried and baked as before.

M3.4.b Prehybridisation

Once DNA was fixed onto filters by baking, the filters were subject to prehybridisation for a minimum of 1 hour (repeated twice for colony blots) in heat sealed plastic bags at the appropriate temperature with shaking.

Prehybridisation solution	
SSC buffer	6x
Denhardt's solution	5x
SDS	0.5% (w/v)
Denhardt's solution (100x)	% (w/v)
BSA	2
Ficoll	2
Polyvinylpyrollidone	2

Poly A^+ DNA was added to a concentration of 50 ug/ml as blocking agent.

M3.4.c Hybridisation

Labelled probe was added after heat denaturation (except in the case of oligonucleotide probes) to the prehybridisation mixture and incubated overnight with shaking at the appropriate temperature. The volume of hybridisation solution was kept to a minimum (1ml/10cm^2 filter) and incorporated radiolabel was added to a concentration of 10^6dpm/ml. Filters were then subject to a washing protocol specific to the

investigation. Filters were then wrapped in cling film and exposed to X-ray film by autoradiography.

M3.5 Sequencing cloned DNA

Sequencing was performed using the Sanger chain termination method in all cases (Sanger et al. 1977). This was modified by two groups both of whose methods were routinely adopted. Fragments of DNA to be sequenced were subcloned into M13mp18 and M13mp19 vectors (Yanish-Perron et al. 1985), transformed into competent E.coli JM101 and plated out in soft agar with recombinant plaques colour selected by the addition of isopropyl-B-D-thiogalactopyranoside and 5-bromo-4-chloro-3-indolyl-B-D-galactopyranoside to the medium as used before (M3.2.g).

M3.5.a Sequencing: Cambridge method

This was performed as directed by Bankier and colleagues of the Medical Research Council Laboratory of Molecular Biology in Cambridge, UK.

Recombinant plaques were picked with Pasteur pipettes and the plug used to inoculate L-broth containing 1% of an overnight grown culture of E.coli JM101. M13 was grown by vigorous shaking at 300 rpm at 37^{o}C for 5-6 hours. Cultures were harvested and 1.5 ml placed in Eppendorf tubes. After centrifugation for 5 minutes, the supernatant was recentrifuged for a further five minutes. Single stranded M13 phage were then precipitated from the supernatant upon the addition of 200ul polyethylene glycol 6000 (20% w/v), 2.5M sodium chloride. Precipitation proceeded at room temperature for 15 minutes. M13 was then collected by centrifugation for 2 minutes and the supernatant completely removed. The

pellet was recentrifuged and any remaining supernatant removed. M13 was resuspended in 100ul TE buffer, and the suspension phenol extracted (without chloroform). DNA was pelleted by ethanol precipitation and resuspended in 32 ul water.

To each of four tubes labelled A,C,G,T was added 2 ul DNA template and 2 ul TM/Primer mixture. These were then incubated at >55°C for 30 minutes.

TM/Primer mixture	ul
Primer (0.2pmol)	1
H_2O	7
TM	1

(TM contains 100mM Tris-HCl (pH 8.5) and 50mM magnesium chloride).

2 ul appropriate NTP mixture was added to the sides of each tube with 2 ul reaction mixture and the tubes centrifuged for 5 seconds to mix the contents.

NTP mixtures (volumes in ul)

	T	C	G	A
0.5mM dTTP	25	500	500	500
0.5mM dCTP	500	25	500	500
0.5mM dGTP	500	500	25	500
10mM ddTTP	50	-	-	-
10mM ddCTP	-	8	-	-
10mM ddGTP	-	-	16	-
10mM ddATP	-	-	-	1
TE	1000	1000	1000	500

Reaction mixture	ul
0.1M dithiothreitol	1
[a-^{35}S]dATP	1
DNA polymerase 1 (Klenow) 1 U/ul	1
H_2O	6

Tubes were incubated at room temperature for 20 minutes then 2 ul dNTP chase was added followed by incubation at 37^{o}C for 15 minutes.

dNTP chase	mM
dTTP	0.5
dATP	0.5
dCTP	0.5
dGTP	0.5

Reactions were terminated by the addition of 2 ul formamide dye mixture.

Dye mixture	
deionised formamide	100ml
xylene cyanol	0.1g
bromophenol blue	0.1g
0.5 M EDTA	2ml

Samples were heat denatured for 5 minutes at 80^{o}C before loading onto polyacrylamide gels. 400 x 150 x 0.4mm gradient gels were run using BRL sharks tooth combs to load samples. Gels contained 8mls 2.5xTBE acrylamide layered with 22 mls 0.5xTBE acrylamide.

	2.5xTBE	0.5xTBE
TBE (10x) (M3.1.e)	250 ml	50 ml
stock acrylamide	150 ml	150 ml
urea	460 g	460 g
bromophenol blue	50 mg	
sucrose	50 g	
H_2O	to 1L	to 1L

Stock acrylamide contained acrylamide (38% w/v), bis-acrylamide (2% w/v). Polymerisation proceeded on the addition of 6 ul ammonium persulphate (10% w/v) and 0.5 ul TEMED per ml of gel. Gels were run at constant 40W until the bromophenol blue dyefront reached the base of the gel. Gels were then fixed in methanol (5% v/v), acetic acid (5% v/v) for 15 minutes then dried under vacuum onto Whatman 3MM paper and exposed to X-ray film.

M3.5.b Sequencing: Sequenase system

Sequenase reactions were followed as according to the manufacturer's instructions. M13 template was prepared in the same way as before, as was the running of samples on gels, only the reaction conditions were different.

To 7 ul template DNA was added 1 ul primer (0.5 pmol) and 2 ul sequencing buffer.

Sequencing buffer	mM
Tris-HCl (pH 7.5)	200
magnesium chloride	100
sodium chloride	250

This was annealed for two minutes at 65°C and cooled slowly to room temperature. To four tubes labelled A, C, G, T, was added 2.5 ul appropriate termination mixture. Each termination mixture contained 80uM all four deoxynucleotides and 8uM one specific dideoxynucleotide, all in 50mM sodium chloride. To the annealed DNA mixture was added:

Reaction mixture	ul
100mM dithiothreitol	1
Labelling mixture	2
[a-^{35}S]dATP	0.5
Sequenase	2

Labelling mixture contained 7.5 uM dGTP, dCTP, dTTP and was diluted 1 in 5 before use. Sequenase was diluted 1 in 8 before use. No information was provided on its unit activity.

The reaction mixture was incubated for 5 minutes at room temperature then 3.5 ul aliquotted to each of the tubes containing dideoxynucleotides at 37°C. After incubation for 5 minutes the reactions were terminated by the addition of 4 ul Stop solution as before. 2 ul samples were analysed by electrophoresis as before after denaturing at 75°C for 2 minutes.

In both sequencing protocols universal primer was used:

5'-GTAAAACGACGGCCAGT-3'

DNA was labelled by deoxyadenosine 5'-[a-^{35}S] thiotriphosphate (Sp isomer), specific activity > 37TBq/mmol.

CHAPTER 3

RESULTS AND DISCUSSION

D1

GENERATING cDNA CLONES

D1.1 Summary of results

cDNA was synthesised from poly A^+ RNA extracted from developing seeds of the castor bean plant Ricinus communis. It was used to generate a cDNA library which was screened for the presence of clones representing the transcript for the precursor to the previously characterised 2S albumin (Sharief and Li, 1982). The probe used was a synthetic oligonucleotide mixture directed to the large subunit of the 2S albumin. This was found to be specific to developing seed cDNA and recognised cDNAs of maximum length 1.1 kb. The probe positively hybridised to 43 plasmids that contained cloned inserts, most of which appeared to be of similar length. Nearly all possessed restriction endonuclease sites in the same places and three clones, 8g8, 10a12 and 14g4, were chosen for further study.

Clone 8g8 recognised a transcript of 1.0-1.1 kb by Northern analysis of developing castor bean seed poly A^+ RNA. It failed to recognise a transcript from germinating seed poly A^+ RNA. All three clones were subcloned into the in vitro transcription vector pGEM 3blue. Transcripts generated from clones 10a12 and 14g4 when placed in an in vitro translation system generated products of apparent molecular weight 34 kDa. Clone 8g8 transcripts failed to translate. These translation products were immunoprecipitable by rabbit polyclonal antibodies that had been raised towards the previously characterised 2S albumin. When translated in the presence of dog pancreatic microsomal membranes,

transcripts from clones 10a12 and 14g4 generated products that were cleaved to an apparent molecular weight of 32 kDa.

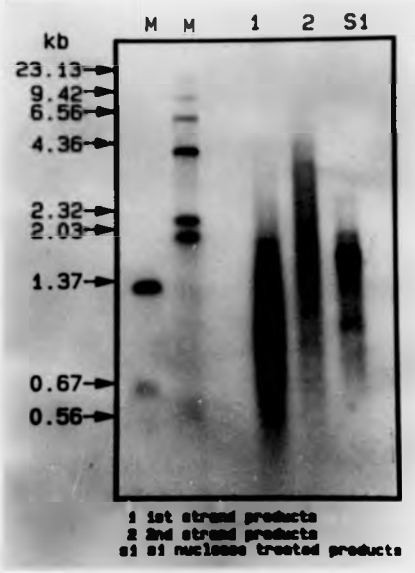
Figure 3

Alkaline agarose gel analysis of the products of cDNA synthesis reactions

Poly A^+ RNA from developing castor bean seeds was subjected to the cDNA synthesis reactions as described (M3.2.a-c). Aliquots of the products of each reaction were analysed on a 1.4% alkaline agarose gel as described (M3.1.e). Samples loaded possessed activity of 20 cpm after ethanol precipitation (M3.1.a). Gels were dried and exposed overnight to film by autoradiography.

Lane 1	First strand cDNA products
Lane 2	Second strand cDNA products
Lane S1	Second strand cDNA products treated with S1 nuclease

Molecular weight markers were the Hind III restriction fragments of lambda DNA as described (M3.1.e) and the Sau 3A restriction fragments of plasmid pBR322. Both sets of molecular weight markers were labelled by the end-fill reaction as described (M3.3.d).


M
M
1
2
S1
kb
23.13
9.42
6.56
4.36
2.32
2.03
1.37
0.67
0.56
1 1st strand products
2 2nd strand products
s1 s1 nuclease treated products

D1.2 RNA extraction and synthesis of complementary DNA

156 ug poly A^+ RNA was generated from 30 g dwarf post-testa castor bean seeds and was found to translate well in a wheatgerm *in vitro* translation system (>360,000 dpm acid precipitable material per ug RNA). This was achieved after many failed extractions, and since tissue was in limiting quantities, practice was achieved by extracting RNA from germinating castor bean seeds. It was found that washing the ethanol precipitated nucleic acid pellet with 3M sodium acetate was a crucial step and this was repeated up to six times. Upon satisfactory *in vitro* translation poly A^+ RNA was used to synthesise complementary DNA (cDNA).

10 ug aliquots of poly A^+ RNA were used in the first strand reaction, which used synthetic oligo-dT to prime the incorporation of deoxynucleotides into DNA by AMV reverse transcriptase. This reaction was typically 30% efficient generating approximately 3 ug single stranded DNA. The second strand synthesis reaction was more efficient (70%) using *E.coli* DNA polymerase 1 and the generated hairpin loop as primer to incorporate deoxynucleotides into DNA, creating 4.2 ug double stranded cDNA. Subsequent use of different enzymes such as the Klenow fragment of *E.coli* DNA polymerase 1, T7 DNA polymerase and AMV reverse transcriptase was found neither to increase the yield nor the length of cDNAs. The hairpin loop created during second strand synthesis was cleaved using S_1 nuclease from *Aspergillus oryzae*. Aliquots of each reaction were run on alkaline agarose gels to diagnose the integrity of the products from each reaction. An example of this is shown in Figure

3. Under the alkaline running conditions DNA migrates as single strands. Hence after second strand synthesis the unit length of cDNAs appeared to have doubled, since both first and second strand products were essentially contiguous, by virtue of the hairpin loop structure. Cleavage of the hairpin loop allowed the products to migrate at unit length again.

The S_1 nuclease method has been thoroughly tried and tested in many laboratories, indeed the ricin and Ricinus communis agglutinin cDNAs were cloned this way in our laboratory (Lamb et al, 1985; Roberts et al, 1985). The method has been largely superseded by the use of RNAse H to generate the conditions for second strand synthesis (Okayama and Berg, 1982). This is due to the fact that S_1 nuclease results in the loss of important 5' information often making the clones incomplete within the coding region (Gubler and Hoffman, 1983). S_1 nuclease was used in this cloning strategy because it was discovered previously that RNAse H would not use castor bean poly A^+ RNA as a substrate, either because the message was nuclease resistant or because an inhibitor was present in the preparation (McGurl, 1986).

Once cDNA demonstrated correct mobility on an alkaline agarose gel it was used in subsequent reactions in readiness for its eventual use in the generation of a cDNA library.

Figure 4

Tailoring of the products of cDNA synthesis reactions

cDNA generated by the reactions described (M3.2.a-c) was subject to the following reactions in a stepwise manner to enable it to generate a cDNA library.

The end-fill reaction was performed as described in section M3.2.e.

The methylase protection reaction was performed as described in section M3.2.f.

Eco RI linkers were radiolabelled according to the kinase reaction as described in section M3.3.a.

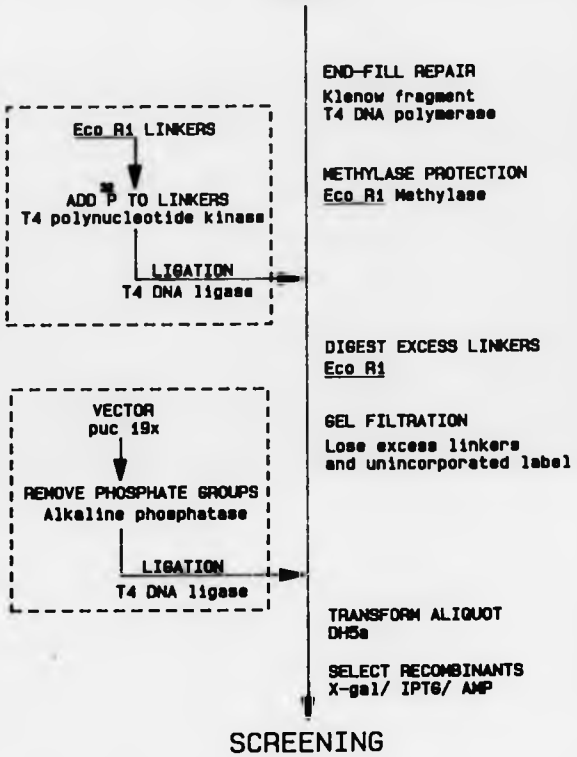
Linkers were then ligated onto cDNA, which was then subject to Eco RI digestion and gel filtration as detailed in section M3.2.g.

Vector puc19x, previously linearised by digestion with Eco RI. was treated with calf intestinal alkaline phosphatase (M3.1.h) and then ligated onto the cDNA (M3.1.i).

DNA was then used to transform cells (M3.1.g).

Bacteria were plated onto a selective medium containing ampicillin, X-gal and IPTG (M3.2.h).

Colonies unable to utilise X-gal as a substrate (remaining white in colour) were individually picked into wells of microtitre plates containing 200 ul selective medium and grown overnight at 37^{o}C. They were then subject to screening (M3.4).


cDNA
END-FILL REPAIR
Klenow fragment
T4 DNA polymerase
Eco R1 LINKERS
METHYLASE PROTECTION
Eco R1 Methylase
ADD 32P TO LINKERS
T4 polynucleotide kinase
LIGATION
T4 DNA ligase
DIGEST EXCESS LINKERS
Eco R1
VECTOR
puc 19x
GEL FILTRATION
Lose excess linkers
and unincorporated label
REMOVE PHOSPHATE GROUPS
Alkaline phosphatase
LIGATION
T4 DNA ligase
TRANSFORM ALIQUOT
DH5a
SELECT RECOMBINANTS
X-gal/ IPTG/ AMP
SCREENING

D1.3 Generation of a cDNA library

After S_1 nuclease digestion to remove the hairpin loop structure, the cDNAs had to be blunt ended in readiness for subsequent ligations. This was achieved by the action of E.coli DNA polymerase 1 Klenow fragment extending recessed 3' ends (Shishodo and Ando, 1981). This represented the first of a series of reactions performed on the cDNA to enable it to generate a library. These are schematically represented in Figure 4.

Following the end-fill repair reaction the cDNA was subject to the action of Eco R1 methylase, to protect internal Eco R1 restriction sites from digestion after the ligation of synthetic Eco R1 linkers to the ends of the cDNAs. This ligation is an inefficient reaction and to ensure addition of linkers to both ends was performed with a molar excess of linkers to cDNA. Excess linkers and digested fragments were removed after Eco R1 digestion by gel filtration. Their removal avoids the production of false positive colonies, which appear to possess a recombinant plasmid, but instead just contain linkers. The cDNA was then ready to be ligated into linearised plasmid vector, with subsequent transformation into competent E.coli. The vector chosen was puc19x, a derivative of puc19 (Yanisch-Perron et al, 1985), produced in this laboratory possessing a Xho I site in the multiple cloning site. The vector, previously linearised with Eco R1, was treated with calf intestinal alkaline phosphatase to remove 5' phosphate residues, thus preventing self-ligation in the subsequent ligation with the cDNA molecules. This phosphatase step was generally successful, but never

completely abolished non-recombinant plasmids from appearing in the library.

After transformation of competent E.coli strain DH5a, cells were plated onto selective medium. Ampicillin selected bacteria transformed with the plasmid vector, whilst recombinant plasmids were selected from colonies appearing white in colour, due to their inability to use the chromogenic substrate X-gal in the presence of the inducer IPTG. Non-recombinant colonies possessing religated vector alone possessed a functional alpha polypeptide from the uninterrupted lac Z gene and so were able to complement the lac Z deletion in the host. With functional beta-galactosidase activity X-gal was utilised and the colonies appeared blue in colour (Davies et al, 1980).

White colonies were individually toothpicked into microtitre wells and grown in selective liquid medium overnight. These were used to inoculate nitrocellulose filters overlaid onto selective solid medium. Colonies on the filters grew overnight and were used for screening. The microtitre liquid cultures were frozen at -70^{0}C after the addition of DMSO to 7% (v/v).

Before screening commenced, colonies were chosen at random and plasmid DNA was prepared. The sizes of the cloned inserts were ascertained by digestion with Eco RI and agarose gel electrophoresis. It became increasingly obvious that some form of size fractionation of the cDNA was required, since all white colonies selected were false positives. Size fractionation of the cDNA prior to vector ligation was attempted on Biogel P-60, Sephacryl S-300, S-400 and S-1000, but electrophoresis and

autoradiography of samples suggested fractionation was not being achieved. cDNA was then electrophoresed in low melting point agarose and size fractionated by removing gel blocks at 100bp intervals. DNA was phenol extracted and utilised as described. This method proved successful and a library of 10,000 colonies was generated using cDNA sized between 0.8 and 1.2 kb.

Figure 5

Synthetic oligonucleotide probe mixture used to screen the cDNA library

The cDNA library was screened by hybridisation to a mixture of 32 synthetic oligonucleotides directed to the large subunit of the previously identified 2S albumin of castor bean (Sharief and Li, 1982).

The amino acid residues chosen to be represented by the oligonucleotides are detailed on the top line of the figure, where N denotes the amino terminal side and C the carboxyl terminal side.

The cysteine residue is numbered 9 in the original sequence determined by Sharief and Li (1982).

The number of codons that could represent each amino acid residue is listed beneath.

The bottom line represents codon sequences, hence the sequence of oligonucleotides synthesised. A base below the third nucleotide of each codon represents the area where degeneracy occurs, hence where either base could be used in the synthesis of an oligonucleotide. To restrict the number of different oligonucleotides to a maximum of 32, leucine was represented by just one codon (CTA) which generated the possibility that the probe may contain two mismatches, and glutamine (residue 16) was represented by only the first two nucleotides of the codons that represent it.

The synthetic 23-mer oligonucleotide mixture was synthesised using an Applied Biosystems 380 B DNA synthesiser and purified by HPLC.

N:	CYS	ASP	HIS	LEU	LYS	GLN	MET	GLN : C
CODONS:	2	2	2	6	2	2	1	2
OLIGO:	TGT C	GAT C	CAT C	CTA	AAA G	CAA G	ATG	CA

Figure 6

Unlabelled cDNA probed with the oligonucleotide mixture

Part A

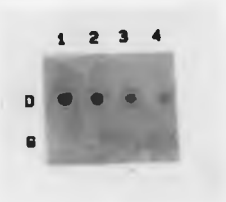
cDNA was generated from both developing (D) and germinating (G) castor bean seed poly A^+ RNA as described (M3.2.a-c) except that no radiolabelled nucleotide was added, all four nucleotides being present in equal amounts. cDNAs were applied to nitrocellulose using a dot blot manifold and screened with the oligonucleotide probe as described (M3.4)

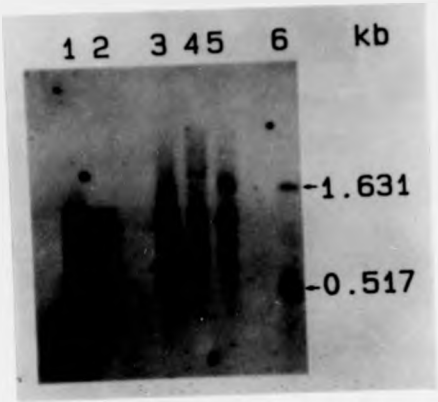
Lane 1 1 ug cDNA
Lane 2 0.5 ug cDNA
Lane 3 0.25 ug cDNA
Lane 4 0.05 ug cDNA

Part B

cDNA was prepared from developing castor bean seed poly A^+ RNA with and without radiolabelled nucleotide and the products run on a 1.4% alkaline agarose gel as described (Figure 3). The DNA fragments were blotted onto nitrocellulose filters (M3.4.a) and screened as above.

Lane 1 2 ug cDNA (unlabelled)
Lane 2 0.5 ug cDNA (unlabelled)
Lane 3 S1 nuclease treated cDNA (labelled)
Lane 4 2nd strand reaction products (labelled)
Lane 5 1st strand reaction products (labelled)
Lane 6 Size markers: pBR322 digested with Hin fI (labelled).


A
1 2 3 4
D
S


B
1 2 3 4 5 6 kb
1.631
0.517

D1.4 Design and specificity of the synthetic oligonucleotide probe

The amino acid sequence of the castor bean 2S albumin had been determined by Sharief and Li (1982). Within the large subunit sequence appeared a stretch of residues which were appropriate for the construction of a mixed synthetic oligonucleotide probe to screen the cDNA library. The details are given in Figure 5. This is a routine method used successfully to isolate clones (Montgomery et al, 1978; Wallace et al, 1981). The probe was longer than the typical 17 bases often used for probing, but remained within the limit of 32 different oligonucleotides.

The redundancy of the genetic code means that most amino acids are encoded by two or four different codon triplets, with three of the most common amino acids being represented by six different codons each. The area chosen in the large subunit possessed a methionine residue represented by only one codon. Near this however was a leucine residue, represented by six codons. Only one codon was chosen, which was the most likely assuming no codon bias in the plant. It still generated the possibility that two mismatches might occur but this was deemed satisfactory, due to the increased length. Single, long synthetic oligonucleotides have been used instead of a mixture of short probes (Jaye et al, 1983; Anderson and Kingston, 1983; Ullrich et al, 1984). The uncertainty at each codon is largely ignored and instead probe length is used to confer specificity. A review of the considerations

over selecting synthetic oligonucleotide probes is given by Lathe (1985).

Ideally, in retrospect, the leucine codon should have been represented by inosine bases which have been used successfully in areas of high degeneracy, since inosine as a base analogue acts as a 'wild-card' binding to all four natural bases at ambiguous positions with or without forming hydrogen bonds (Takahashi et al. 1985).

The possibility of two mismatches being present within the closest matching oligonucleotide resulted in hybridisation temperatures being reduced to 37°C. Washing of probed filters took place initially at 37°C and was monitored at 5°C intervals to 52°C in 6xSSC, 0.1% SDS (w/v).

The first filters to be probed were those shown in Figure 6. Part A shows that the oligonucleotide mixture hybridises to cDNA made only from poly A^+ RNA taken from developing castor bean seeds, and not to cDNA made from germinating seed poly A^+ RNA. The fact that the probe appeared to be developmentally specific to the stage where the 2S albumin was being deposited, and was specific so as not to cause background binding meant that it seemed suitable for screening the cDNA library. Part B shows that in lanes 1 and 2 the oligonucleotide probe recognises a smear of cDNA products with greatest length 1.1 kb. This length seemed to be suitable for a clone representing a transcript that generates a 34 kDa precursor. The smear seen could represent cDNA that was degraded by nucleases, or reflect the poor integrity of the poly A^+ RNA template, or be a reflection of the poor efficiency of the AMV reverse transcriptase in generating the first strand reaction products (Glover, 1985).

Whatever the cause of the smearing it was decided that to avoid too many incomplete clones being isolated, the library should be constructed from cDNA size fractionated between 0.8-1.2 kb, which was achieved as described (01.3).

Figure 7

Screening the cDNA library

Colonies were grown overnight in microtitre plate wells with selective medium then transferred to nitrocellulose overlaid on selective solid medium, grown overnight and prepared on nitrocellulose as described (M3.4.a). After prehybridisation (M3.4.b), the filters were hybridised overnight (M3.4.c) with labelled oligonucleotide probe (M3.3.a and figure 5) at 37°C. The filters were then washed in 2xSSC buffer (M3.4.a), 0.1% (w/v) SDS at successively higher temperatures (37°C, 42°C, 47°C), with autoradiography performed between each wash. The figure indicates film developed for each wash temperature after exposure overnight at -70°C. One microtitre tray is represented, with two control wells located top left, one with no bacterial growth to confirm sterility and one containing bacteria transformed with plasmid puc19x.

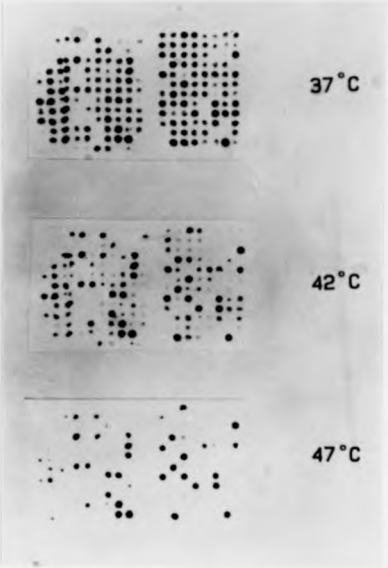

37°C
42°C
47°C

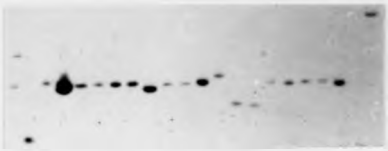
Figure 8

Southern blot of plasmid DNA extracted from colonies that hybridised to the oligonucleotide probe on primary screening

Plasmid DNA was extracted (M3.1.c) from the cells of colonies that hybridised strongly to the oligonucleotide probe. 1 ug plasmid DNA from each extraction was digested with Eco RI restriction endonuclease (M3.1.f), then electrophoresed on 1.5% agarose gels buffered with TBE, visualised and photographed (M3.1.e). The gels were subsequently blotted onto nitrocellulose filters and probed with the labelled oligonucleotide mixture as described (M3.4).

Parts A1 and B1 of the figure refer to photographs of the gels taken before blotting, whilst parts A2 and B2 refer to photographs of the autoradiographs that were exposed to the probed filters overnight at -70^{0}C. Each gel was loaded from left to right with 20 (gel A) or 23 (gel B) different plasmids, 1 ug Eco RI digested puc19x and 2 ug lambda DNA digested with Eco RI and Hind III.

A1



A2

B1

B2

Table 1

Classification of cDNA clones by restriction endonuclease digest

38 plasmids that positively hybridised to the oligonucleotide probe were subject to restriction endonuclease digestion by enzymes Pvu II, Sac I and Sph I. Digested plasmid DNA was then electrophoresed on a 1.5 % agarose gel buffered with TBE and DNA fragments visualised (M3.1.e). Clones were classified according to the number of restriction sites located within the cloned fragment, excluding those in the vector.

CLONES	GROUP	Pvu II	Sph I	Sac I
27	1	2	1	2
5	2	2	1	3
2	3	2	0	2/3
1	4	2	0	0
1	5	DOUBLE INSERTS		
1	6			
1	7	3	1	3

D1.5 Selection of positively hybridising clones

The cDNA library was screened with the oligonucleotide probe and the filters washed as described (Figure 7). Successively higher temperatures caused non-specifically bound probe to be removed leaving clearer signals from clones. DNA from colonies containing vector puc19x did not hybridise to the probe. In total 67 colonies hybridised to the oligonucleotide probe. Figure 8 shows Southern blots of plasmids, extracted from those colonies, that yielded a fragment on digestion with Eco R1 (43 did so, 24 were linearised). The Southern blots were probed with the oligonucleotide mixture and washed at 47°C. Most plasmids generated a single insert which hybridised strongly to the probe. Inserts appeared to be of similar lengths, approximately 1.0 kb, others ranged from 1.1 kb to 0.5 kb, which was surprising considering the size fractionation procedure adopted.

The different clones were subjected to restriction enzyme digestion by common enzymes and it was discovered that most were cut by Pvu II, Sph I and Sac I. On this basis, the clones were categorised according to the number of restriction sites located within the insert. The results are listed in Table 1. It would seem that the large majority belonged to group 1, from which was chosen clones 8g8 and 10a12, with the second biggest group possessing an extra Sac I site, from which was chosen clone 14g4 for further study. Clone 14g4 was approximately 100 bp longer at the 3' end, where the extra Sac I site was located, than the group 1 clones (see page 189). Groups 3-7 possessed clones which seemed to

resemble group 1 and 2 members but appeared anomalous in terms of their numbers represented. Groups 5 and 6 contained members that were suspected to be two cDNA fragments ligated together before the ligation of the synthetic linkers. In view of the numbers concerned groups 3-7 were not subject to further examination. It was believed that the number of types of clones isolated was a reflection of the steady state transcript level of the albumin gene(s).

Nomenclature

Clones are referred to by their microtitre grid reference. 8g8 refers to the DNA insert only. The construction of insert and pUC19x led to the formation of the pSI plasmids, for example pSI8g8.

Figure 9

Northern blot of developing castor bean poly A^+ RNA probed with cDNA clone 8g8

Poly A^+ RNA extracted from developing and germinating castor bean seeds was electrophoresed on a 1.3% (w/v) agarose /formaldehyde gel buffered with MOPS (M2.3.a). The gel was blotted onto nitrocellulose and probed (M2.4/M3.4) with clone 8g8 DNA labelled by nick-translation (M3.3.b). The filter was washed in 6xSSC, 0.1 % (w/v) SDS at 42^oC for 30 minutes, then in 2xSSC, 0.1 % (w/v) SDS at 42^oC for 30 minutes then twice in 0.1xSSC, 0.1 % (w/v) SDS at room temperature for 15 minutes after which the filter was exposed to film overnight at -70^oC. Size markers were visualised on the filter by subsequent staining with methylene blue (M2.4).

Lane 1	2 ug developing castor bean poly A^+ RNA
Lane 2	1 ug developing castor bean poly A^+ RNA
Lane 3	0.5 ug developing castor bean poly A^+ RNA
Lane 4	0.1 ug developing castor bean poly A^+ RNA
Lane 5	1 ug germinating castor bean poly A^+ RNA

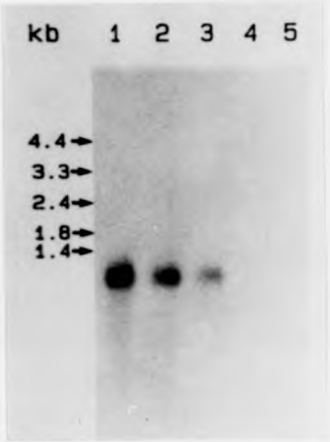
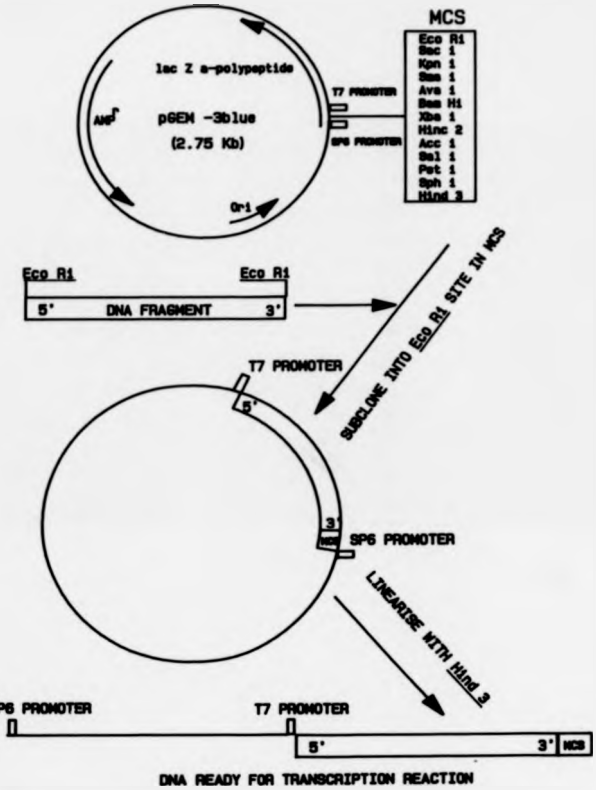

kb
1
2
3
4
5
4.4
3.3
2.4
1.8
1.4

Figure 10

Schematic representation of the subcloning of cDNA clones into plasmid pGEM 3blue and preparation of the plasmid for in vitro transcription

Plasmid pGEM 3blue was linearised by restriction endonuclease digestion with Eco RI. the site of cleavage being located within the multiple cloning site (MCS). The MCS is located between the SP6 and T7 promoter regions which oppose each other. It is also located at the start of the coding region for the lac Z alpha-polypeptide, thus enabling subsequent colour selection. Cloned cDNA was excised from vector puc19x by restriction endonuclease digestion with Eco RI and isolated from an agarose gel (M3.1.j). This was ligated into pGEM 3blue using T4 DNA ligase as described (M3.1.i). E.coli was transformed with the ligation (M3.1.g) and plated onto solid medium containing ampicillin, IPTG and X-gal (M3.2.h). Plasmid DNA was prepared from those bacteria containing recombinant plasmids with the 5' end of the subcloned fragment proximal to the T7 promoter region (M3.1.d). Construct DNA was then restriction endonuclease digested with Hind III, an enzyme which cuts within the MCS located downstream of the subcloned DNA but which does not cut within the subcloned fragment. The DNA was then ready for the transcription reaction as described (M3.1.l).



DNA READY FOR TRANSCRIPTION REACTION

Figure 11

cDNA clones in vitro transcribed, translated, and immunoprecipitated with antibodies raised towards the previously characterised 2S albumin.

cDNA clones 14g4 and 10a12 were subcloned into vector pGEM 3blue as described (M3.1.1 and Figure 10). In vitro transcripts were generated and used within a wheatgerm in vitro translation reaction (M2.2.b). Of the 25 ul translation reaction, 12.5 ul were subject to immunoprecipitation with rabbit antibodies raised towards the previously characterised 2S albumin (M2.2.d). Translational products and immunoprecipitations were resolved on a 15% (w/v) denaturing polyacrylamide gel under reducing conditions. The gel was then fluorographed, dried and exposed to film by autoradiography overnight at -70^{o}C (M2.2.e).

Lane 1 12.5 ul translation products from clone 10a12 transcripts.

Lane 2 12.5 ul translation products from clone 10a12 transcripts, immunoprecipitated.

Lane 3 12.5 ul translation products from clone 14g4 transcripts.

Lane 4 12.5 ul translation products from clone 14g4 transcripts, immunoprecipitated.


1
2
3
4
kDa
92
69
46
30
24
14

Figure 12

cDNA clones 10a12 and 14g4 in vitro transcribed and in vitro translated within a rabbit reticulocyte lysate system in the presence and absence of dog pancreatic microsomal membranes.

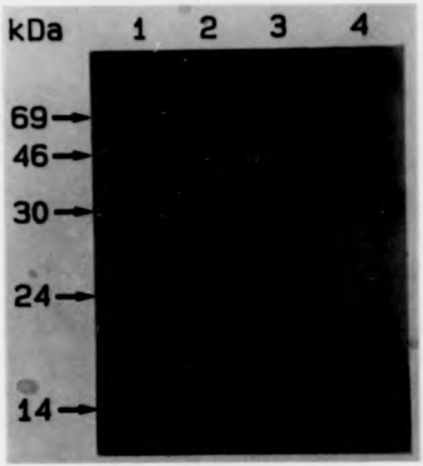
Transcripts from clones 10a12 and 14g4 were generated as previously described (M3.1.1 and Figure 10). Rabbit reticulocyte lysate in vitro translations were performed using the transcripts as substrate. Translation reactions were performed in both the presence and absence of dog pancreatic microsomal membranes as described (M2.2.a). Translation products were resolved on a 15% (w/v) denaturing polyacrylamide gel, which was subsequently fluorographed, dried and exposed to film overnight by autoradiography at -70^{o}C.

Lane 1 Translational products from clone 10a12 transcripts, in the absence of membranes.

Lane 2 Translational products from clone 10a12 transcripts, in the presence of membranes.

Lane 3 Translational products from clone 14g4 transcripts, in the absence of membranes.

Lane 4 Translational products from clone 14g4 transcripts, in the presence of membranes.


kDa
1
2
3
4
69
46
30
24
14

D1.6 Further identification of the positively hybridising clones

Having shown that the synthetic oligonucleotide probe recognised the cloned inserts it was necessary to provide further positive identification before embarking on a sequencing programme. The clones chosen were obviously related by virtue of their similar restriction maps, so clone 8g8 was used to probe a Northern blot of the poly A^+ RNA used to generate the cDNA, as well as germinating poly A^+ RNA. The result is shown in Figure 9. The result reflects what was seen for the probing of the developing cDNA blot with the oligonucleotide probe (Figure 6, parts A and B), that the clone is developmentally regulated and hybridises to a transcript of size 1.0-1.1 kb. This blot suggests also that the poly A^+ RNA was not heavily degraded as was suggested (D1.4). The clones appear to be near full length. The lack of a transcript in germinating poly A^+ RNA is not surprising since the storage proteins are deposited during seed development only and mobilised during germination (Kermode et al, 1985; Youle and Huang, 1978a; Gifford et al, 1983; Butterworth and Lord, 1983).

That the clone was developmentally specific and hybridised to a synthetic oligonucleotide directed to the large subunit of the previously characterised 2S albumin did not yet positively identify the clone as the precursor to the 2S albumin. Since only one castor bean 2S albumin has so far been characterised (Sharief and Li, 1982), although at least two have been reported (Tully and Beevers, 1976), there exists the possibility that two or more similarly related 2S albumin precursors

may exist and be recognised by the probe. Also, since the probe was directed to an area of high conservation amongst the storage protein superfamily (region B, Kreis *et al*, 1985b) it seemed as likely that the probe might recognise cDNAs representing not only albumins but other storage proteins as well. A more positive identification was needed. Studies on the precursor generated from poly A^+ RNA by *in vitro* translations had shown that it could be immunoprecipitated by antibodies directed towards the 2S albumin (McGurl, 1986). It had also been demonstrated that the precursor possessed an N-terminal signal sequence which could be removed by performing rabbit reticulocyte lysate *in vitro* translations in the presence of dog pancreatic microsomal membranes (Butterworth and Lord, 1983). This latter observation was fully in line with the precursors for the napin of *Brassica napus* (Crouch *et al*, 1983), the sulphur rich 10 kDa prolamin of rice seeds (Masumura *et al*, 1989), the pea sulphur rich albumin (Higgins *et al*, 1986), the sunflower albumin (Allen *et al*, 1987), and the Brazil nut 2S albumin (Altenbach *et al*, 1987) which all demonstrate or predict signal sequences that are cotranslationally removed.

It was decided that to see if polypeptides generated from the cDNA clones could mimic the action of the precursor polypeptide by being immunoprecipitated by 2S albumin antibodies and by being cleaved of a signal sequence. The cloned inserts were subcloned into the *in vitro* transcription vector pGEM 3blue (Figure 10), from which transcripts were generated. These were utilised in a wheatgerm *in vitro* translation and the products immunoprecipitated by 2S albumin antibodies. Figure 11 shows that the products of *in vitro* translation were recognised by the

antibodies. In a simultaneous experiment the antibodies immunoprecipitated the 2S albumin precursor generated from an *in vitro* translation of developing poly A^+ RNA from castor bean seeds used to generate the cDNA. The antibodies failed to recognise any other product. It is known that the precursor to the 2S albumin is less reactive to antibodies raised towards the 2S albumin when the signal sequence is still present (Roberts and Lord, 1981b). This probably results from the N-terminally cleaved precursor conformation more closely resembling that of the mature protein than the uncleaved precursor conformation (McGurl, 1986). It is thus not surprising that the immunoprecipitation did not generate a stronger result (Figure 11, lanes 2 and 4). Clones 10a12 and 14g4 transcripts translated well, whilst transcripts from clone 8g8 failed to translate. This was a reproducible result. The fact that clones 10a12 and 14g4 transcripts translated and generated polypeptide products that migrated on reducing polyacrylamide gels with a molecular weight similar to that generated from the native poly A^+ transcript (apparent molecular weight 34 kDa) suggested that these two clones were full length, that is, containing full open reading frames within their sequences. That the products were also immunoprecipitable seemed further to suggest that the clones did indeed represent the 2S albumin precursor.

Clone 10a12 transcripts appeared to generate a product that was slightly smaller than the product from clone 14g4. It seemed possible then that clone 10a12 represented a different member of the 2S albumin precursor family. This would be backed up by the fact that clone 10a12 was classified as a group 1 clone whilst clone 14g4 was a group 2 clone. The

2S albumins are known to be a heterogeneous group. Sharief and Li (1982) when directly sequencing the subunits of the characterised 2S albumin reported heterogeneity at one position in the large subunit, although this will be shown later to be an incorrect observation (D2.c). It has also been noted in our laboratory that 2-D gel analysis of the immunoprecipitates of the products from in vitro translation of developing castor bean message using antibodies raised towards the previously characterised 2S albumin shows a cluster of different members, suggesting many isoforms (Lord, Warwick, UK). This type of observation has been seen with other 2S albumin precursors (Laroche-Reynal and Delseny, 1986).

When transcripts from clones 10a12 and 14g4 were utilised in a rabbit reticulocyte lysate in vitro translation in the presence and absence of membranes, it was noted that a signal sequence appeared to be cleaved generating products of apparent molecular weight 32 kDa (Figure 12, lanes 2 and 4). The cleavage of the signal sequence appeared to be complete with the product of clone 14g4 (Figure 12, lanes 3 and 4) whilst the reaction was not so efficient with clone 10a12 (Figure 12, lanes 1 and 2). This again was a reproducible result suggesting a difference between the two clones. Once cleaved however, both products appeared to have the same molecular weight suggesting that differences previously observed between them occurred only within the signal sequence. As will be seen later differences were found between clone 10a12 and 14g4 in the signal sequence (Figure 18).

The clones behaved as predicted for those representing the 2S albumin precursor and it was decided to sequence them to provide the final

clarification of their identity. If an open reading frame were to be located it ought to contain the sequence of the subunits of the previously characterised 2S albumin.

D1.7 Conclusions

The cDNA library generated from developing castor bean seeds produced clones that hybridised to a probe directed to the large subunit of the previously characterised 2S albumin. Two of the clones studied appeared to be full length and behaved in a similar way to the native 2S albumin precursor when *in vitro* transcribed and translated. Recognition by 2S albumin antibodies and signal sequence cleavage strongly suggests that the precursor to the 2S albumin has been cloned. True identification will be possible upon verification of the cDNA sequence.

D2 cDNA CLONES: SEQUENCING AND ANALYSIS

D2.1 Summary of results

The identity of the clone 8g8 was confirmed as the precursor to the previously characterised 2S albumin (Sharief and Li, 1982) by sequencing an internal 560 bp restriction fragment. The entire sequence was then determined for clones 8g8 and 14g4. Clone 14g4 was 1097 bp in length whilst clone 8g8 was 948 bp. An open reading frame (ORF) of 258 residues, with molecular weight 29,329 Da, was located wholly within clone 14g4. This ORF contained the entire 2S albumin sequence at the carboxyl end, with the small subunit separated from the large subunit by a three amino acid peptide linker. The small subunit was exactly identical to the published sequence, whilst the large subunit differed slightly in several respects these being explained by faults in the previously published sequence. Clone 8g8 was smaller than clone 14g4 at both the 5' and 3' ends, resulting in an incomplete ORF at the amino terminus. However, both clones sequenced possessed polyadenylated tails at the 3' ends suggesting that clone 8g8 was naturally smaller here than clone 14g4. A potential hairpin loop was discovered in association with the first of three polyadenylation consensus sequences found in the 3' untranslated region. Clone 8g8 also differed from clone 14g4 at the 5' end within the first 18 bases, but was then exactly identical throughout the sequence except for one base mismatch resulting in the substitution of proline for threonine in the ORF.

Secondary structure, hydrophobicity and antigenicity were predicted for the ORF. A signal sequence and putative cleavage site were predicted

also for the ORF from hydrophobicity studies, consensus sequences and homologies with other 2S albumin signal sequences. Also contained within the ORF between the signal sequence and the previously characterised 2S albumin were regions of high glutamine and cysteine residues, which shared homology with other 2S albumin large and small subunits from a wide variety of plant sources.

Figure 13

Southern blot of different restriction endonuclease digests of plasmids pSI8g8 and pSI10a12 probed with the oligonucleotide mixture.

1 ug aliquots of plasmids pSI10a12 and pSI8g8 were subject to digestion with one or more of the following restriction enzymes: Sac I; Pvu II; Eco RI. Fragments were resolved on a 2% (w/v) agarose gel buffered with TBE. After visualisation and photography the DNA fragments were transferred to nitrocellulose filters by Southern blotting (M3.4.a) and were probed with the labelled oligonucleotide mixture as before (M3.4 and Figure 5).

Part A of the figure shows the restriction fragments generated by digestion whilst part B shows the autoradiograph after exposure to the filter overnight at -70°C. Lanes 1-6 refer to digestions performed with plasmid pSI8g8 whilst lanes 7-12 refer to digestions performed with plasmid pSI10a12.

Lanes 1,7	Digestion with Sac I
Lanes 2,8	Digestion with Pvu II
Lanes 3,9	Digestion with Sac I and Pvu II
Lanes 4,10	Digestion with Sac I and Eco RI
Lanes 5,11	Digestion with Pvu II and Eco RI
Lanes 6,12	Digestion with Sac I, Pvu II and Eco RI
Lanes 13	Size markers: 2ug lambda DNA cut with Hind III

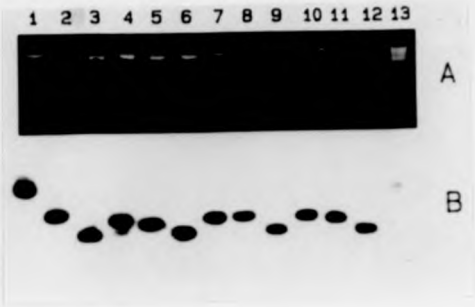
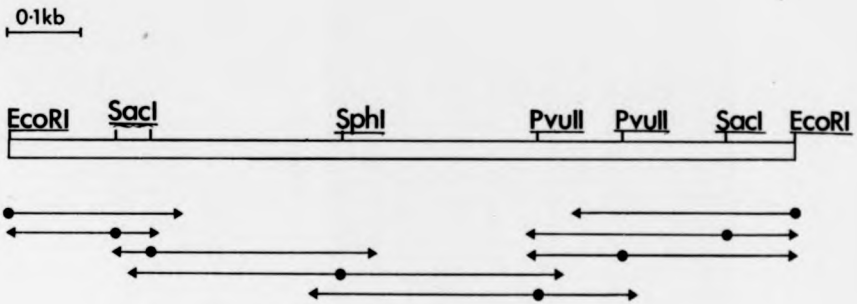

1 2 3 4 5 6 7 8 9 10 11 12 13
A
B

Figure 14

Restriction map and sequencing strategy for clone 14g4.

Clone 14g4 is shown with restriction endonuclease sites above and the sequencing strategy adopted below. Filled circles represent the site from which sequencing was started and the arrow represents the direction and extent of sequence generated.

Facing -174-


0·1kb
EcoRI
SacI
SphI
PvuII
PvuII
SacI
EcoRI

D2.2 Positive identification of cDNA clones and sequencing strategy

It was decided that rather than sequence an entire cDNA clone to positively identify it as the precursor to the 2S albumin, it would be better to analyse the area of the clone directly containing the sequence encoding the mature 2S albumin. Since the mature protein was known to possess a molecular weight of 11 kDa, whilst the precursor was of apparent molecular weight 34 kDa it seemed likely that an internal fragment would need to be subcloned and sequenced in order to achieve this. The oligonucleotide probe was directed towards the mature large subunit sequence, so it was decided to digest the clones with restriction endonucleases compatible with the multiple cloning site of the sequencing vector M13mp19 (Yanisch-Perron *et al.* 1985), and to see which of the smallest fragments would hybridise to the oligonucleotide. The result of this is shown in Figure 13.

The smallest fragment generated was of 560 bp representing an internal *Sac I/Pvu II* fragment. With 300 bp of sequence information being readily obtainable by the dideoxynucleotide chain termination method of Sanger *et al* (1977), both ends of the fragment could be sequenced and overlapped providing information on the area to which the oligonucleotide probe had hybridised. *Pvu II* provides blunt ended molecules, so the internal fragment of clone 8g8 was force-cloned into both the *Sac I/Sma I* cut M13mp18 and M13mp19 double stranded replicative forms.

The fragment was sequenced and an ORF was discovered that contained the small subunit and part of the large subunit of the characterised 2S albumin. Upon this information the restriction maps of 8g8 and 14g4 were identified and constructs of M13 generated containing restriction fragments of the clones, such that both strands of each clone could be sequenced with overlapping guaranteed. Both strands of both clones were sequenced twice. The sequencing strategy for clone 14g4 is given in Figure 14. A similar strategy was adopted for clone 8g8 except that the 3' Sac I site was absent due to it being a smaller clone.

Figure 15

The nucleotide sequence of DNA clones complementary to 2S albumin precursor mRNA and the amino acids deduced from this sequence.

The nucleotide and derived amino acid sequence of 14g4 is shown. The differences found in the 8g8 sequences are indicated where they occur, otherwise the sequences are identical. The entire nucleotide sequence of each clone is delineated between square brackets. In the 3' untranslated region potential polyadenylation signals are indicated by underlining and a hairpin loop is indicated in bold type. The small and large polypeptide sequences of the previously characterised castor bean 2S albumin are boxed and coloured green and yellow respectively. Numbering of the amino acid sequence is based upon the proposed site of signal peptide cleavage.

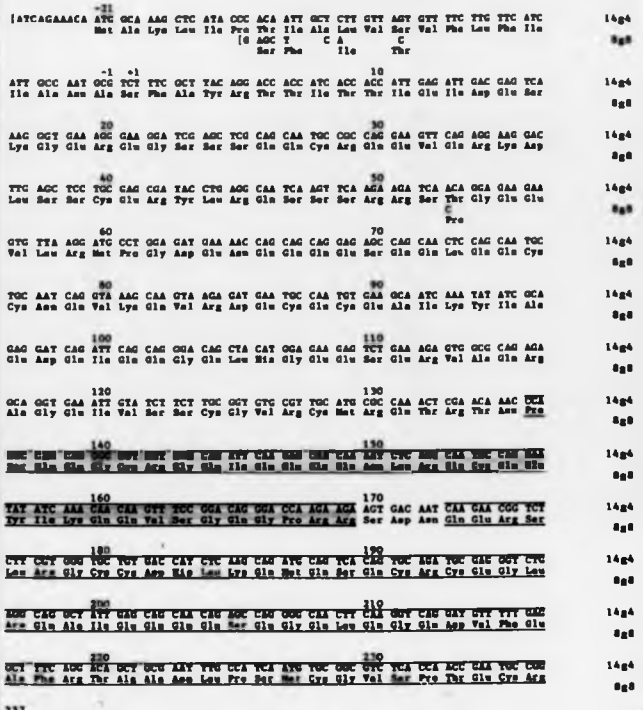


Figure 16

Comparison of the derived amino acid sequence obtained from clone 14g4 with the primary structure of the previously characterised castor bean 2S albumin

The open reading frame generated from the nucleotide sequence of clone 14g4 (Figure 15) was compared with the amino acid sequence generated directly by automated Edman degradation of the castor bean 2S albumin by Sharief and Li (1982).

High homology was found at the carboxyl terminus of the derived amino acid sequence. The upper line of each double row in the figure indicates the derived sequence from clone 14g4 which is only shown from residues 157-258, where residue 1 is the first residue of the open reading frame. The lower line indicates the previously characterised 2S albumin sequence, with small subunit residues numbered 1-34 and large subunit residues numbered 35-95. Gaps have been introduced to maximise homology. Where dashes are present between the two sequences, identical amino acid comparisons have been made.

Clone 8g8 generated an identical sequence to clone 14g4 in this region.

```
157         ProSerGlnGlnGlyCysArgGlyGlnIleGlnGluGlnGlnAsnLeuArgGlnCysGlu
              |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
1           ProSerGlnGlnGlyCysArgGlyGlnIleGlnGluGlnGlnAsnLeuArgGlnCysGln

177         GluTyrIleLysGlnGlnValSerGlyGlnGlyProArgArgSerAspAsnGlnGluArg
              |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
21          GluTyrIleLysGlnGlnValSerGlyGlnGlyProArgArg          GlnGluArg

197         SerLeuArgGlyCysCysAspHisLeuLysGlnMetGlnSerGlnCysArgCysGluGly
              |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
38          SerLeuArgGlyCysCysAspHisLeuLysGlnMetGlnSerGlnCysArgCysGluGly

217         LeuArgGlnAlaIleGluGlnGlnGlnSerGlnGlyGlnLeuGlnGlyGlnAspValPhe
              |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
58          LeuArgGlnAlaIle   GlnGlnGln          GlnLeuGlnGlyGlnAsnValPhe

237         GluAlaPheArgThrAlaAlaAsnLeuProSerMetCysGlyValSerProThrGluCys
              |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
74          GluAlaPheArgThrAlaAlaAsnLeuProSerMetCysGlyValSerProThrGlnCys


257         ArgPhe
              |  |
94          ArgPhe
```

Matches = 93 Mismatches = 2 Unmatched = 7
Length = 102 Matches/length = 91.2 percent

Figure 17

Hairpin loop present at the 3' nontranslated region of clones 8g8 and 14g4.

A sequence enabling a hairpin loop structure was discovered beginning at base 819 of clone 14g4 and was also present in clone 8g8. Immediately upstream of this was located a sequence (underlined) that resembled the polyadenylation signal consensus sequence found in eukaryotic sequences (Proudfoot and Brownlee, 1976).

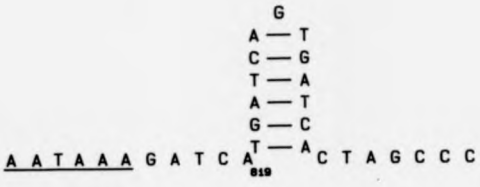

G
A — T
C — G
T — A
A — T
G — C
T — A
A A T A A A G A T C A C T A G C C C
619

Figure 18

A comparison of the castor bean albumin precursor predicted signal sequences with those of other plant albumins

The incomplete signal sequence of 8g8 and the complete sequences of 10a12 and 14g4 are compared with the sunflower albumin (Allen et al. 1987), the napin of Brassica napus (Crouch et al. 1983) and the Brazil nut sulphur-rich albumin (Altenbach et al. 1987).

Sequences are aligned to maximise homology, shown in boxed, yellow areas. Comparisons amongst the other plant 2S albumins are not detailed. Standard one letter code for amino acids is used.


8g8
10a12
14g4
Sunflower albumin
Napin
Brazil nut albumin

D2.3 Sequence data generated from the clones 8g8 and 14g4

The open reading frame

The entire nucleotide sequence of clone 14g4 was established and is shown in Figure 15. The sequence excludes the synthetic Eco RI linkers situated at the ends of the sequence. It is 1097 bases long including a polyadenylated tail of twelve bases. When the sequence was translated in all six reading frames an open reading frame (ORF) of 258 residues was generated which had a calculated molecular weight of 29,329 Da. Whilst this does not match the apparent molecular weight of 34 kDa as seen with the native precursor, it has already been shown that under *in vitro* transcription/translation conditions, the clone produces a polypeptide that migrates on polyacrylamide gels with the apparent molecular weight of 34kDa (Figure 11). It was shown that under non-reducing conditions the precursor migrated as a 22 kDa polypeptide. This was believed to be due to the high cysteine content contained within the precursor (McGurl, 1986). This would appear to be the case, since the ORF contains 16 cysteine residues out of a total of 258 (6.2%). This anomalous behaviour has been reported in other high cysteine content precursors of storage proteins (Allen *et al*, 1987).

The signal sequence

The ORF, known already to be full length (D1.6), was found to possess an ATG codon at the 5' end which is considered to be the initiation codon for translation. Following this are a stretch of hydrophobic residues which appear to represent a putative signal sequence, necessary to

direct the prepropeptide to the endoplasmic reticulum. It is known from both *in vivo* and *in vitro* studies that the precursor to the 2S albumin is transported into the endoplasmic reticulum and a signal peptide is cotranslationally removed (Lord, 1985a; McGurl, 1986). This was further confirmed when a signal sequence appeared to be cleaved whilst *in vitro* generated transcripts from clones 14g4 and 10a12 subcloned into an *in vitro* transcription vector, were translated with dog pancreatic microsomal membranes in a rabbit reticulocyte lysate system (Figure 12). According to the prediction methods of Watson (1984) and von Heijne (1983, 1984, 1985) the ORF has a classical signal sequence in that three domains can be located: a positively charged N-terminal region occurs with a lysine residue at position -19 (Figure 15), followed by a core hydrophobic domain spanning from leucine at position -18 to alanine at -3, leading to the more polar carboxyl terminus containing an asparagine at -2. This then defines the site of signal cleavage after 21 residues between the alanine and serine (numbered on Figure 15 as -1 and +1 respectively). This is in line with the -3,-1 design (von Heijne, 1983), with alanine at both these positions. The ORF displayed in Figure 15 is numbered from the site of signal cleavage, with the first residue of the propeptide, serine, labelled as +1. Unless otherwise stated all residues quoted are numbered according to this convention.

The signal sequences for two other clones 8g8 and 10a12 were determined from the ORFs generated from the entire sequence of clone 8g8 and the 5' end of clone 10a12. These were compared with the signal sequences determined for other 2S albumins (Figure 18). Homologies were found between all three clones with clone 10a12 more homologous to clone 14g4

than 8g8. The signal sequence of 10a12 was shorter than 14g4 by two residues. This was anticipated in the *in vitro* transcription/translation results, whilst the preproalbumins generated from clones 10a12 and 14g4 were of slightly different size, the proalbumins generated after signal cleavage were not (Figures 11 and 12). Clone 8g8 is seen to be devoid of an amino terminal methionine. This would explains the inability of its *in vitro* generated transcripts to translate in an *in vitro* system. In comparison with the other plant 2S albumin signal sequences limited homology was found. All except the *Brassica* napin possessed alanine residues at positions -1 and -3. Clone 14g4 in common with the other plant albumin signals compared possesses an alanine residue after the amino terminal methionine.

That the three clones analysed possessed different signal sequences, yet as will be seen later showed an almost absolute homology within the proalbumin sequence comes as no surprise. A study by Lazaro *et al* (1988) into the signal peptide homology of the sweet protein thaumatin II and unrelated cereal alpha-amylase/trypsin inhibitors suggested that signal peptides seemed to be diverging at a faster rate than the mature sequences within the alpha-amylase/ trypsin inhibitor family. This is credible since the pressure on a signal peptide sequence to remain constant should be less than on the structurally defined mature protein, as long as the divergence in the signal sequence does not prevent it from performing its function. Hence the differences seen between clone 8g8 and 14g4 signal sequences appear to be conserved ones with changes between leucine and isoleucine, serine and threonine and in the case of phenylalanine and isoleucine the hydrophobicity is conserved.

Identification of the 2S albumin

A sequence corresponding to the small subunit of the characterised 2S albumin (Sharief and Li, 1982), was found between residues 136 and 169, and is boxed in green in Figure 15. This generates a polypeptide of molecular weight 3,957 Da. This was followed by 3 amino acid residues, not located within the mature protein, followed by a sequence of amino acids (residues 173 to 237) which closely relates to the published large subunit sequence. This is boxed in yellow in Figure 15 and generates a polypeptide of molecular weight 7,349 Da. When these regions of the ORF were compared with the previously published sequence several discrepancies were noted (Figure 16). Whilst the ORF small subunit possessed exact identity with the published sequence the large subunit did not. Residue 213 (234 in Figure 16) is aspartate rather than asparagine and residue 234 (residue 255 in Figure 16) is glutamate rather than glutamine. More significantly, the large subunit sequence derived from the ORF is 4 residues longer than the published sequence. This discrepancy is readily explained. In the published sequence, Sharief and Li found both serine and leucine at position 33 of the large subunit (residue 67 in Figure 16), followed by GLN-GLY-GLN at the carboxyl terminus of a fragment generated during sequencing. This was explained as heterogeneity of the sample. In fact it seems as though they generated two fragments which were sequenced simultaneously, since the ORF sequence in this region (residues 205-212 or residues 226-233 in Figure 16) is SER-GLN-GLY-GLN-LEU-GLN-GLY-GLN. At the end of the derived large subunit sequence is a stop codon TAA which terminates translation.

Processing of the linker peptide and the 2S albumin subunits from the preproalbumin

The region between the two subunits in the ORF represents a linker peptide which must be post-translationally processed to separate the two subunits. Many other 2S albumin precursors show this relationship between the subunits, with the smaller subunit separated from the larger subunit by a linker peptide, the larger subunit being positioned at the carboxyl terminus of the prepropeptide: in the prepronapin of Brassica napus, one of the most well defined processed precursors, a linker is removed between residues 74 and 94, the small subunit being liberated by another cleavage step at position 38, which is post-translational and distinct from the removal of the signal sequence which is performed cotranslationally (Crouch et al, 1983); the Brazil nut methionine-rich albumin is proteolytically processed in a similar way (Altenbach et al, 1987; Sun et al, 1987). Some storage proteins other than 2S albumins have also been seen to arise from a single prepropeptide, following cotranslational cleavage of a signal peptide and post-translational endoproteolytic cleavage of the propeptide to generate mature subunits within the protein bodies. These include the legumin and vicilin type storage proteins and certain lectins such as those from Vicia faba, Oryza sativa and Pisum sativum (Chrispeels et al, 1982ab; Nielson, 1984; Croy et al, 1980; Spencer and Higgins, 1980; Higgins and Spencer, 1981; Hemperly et al, 1982; Stinissen et al, 1983; Higgins et al, 1983ab and Spencer et al, 1983).

Within the castor bean itself it is known that the A and B chains of ricin, located with the 2S albumins in the soluble matrix of the protein

bodies, are derived from a preproricin and are liberated after removal of a 12 amino acid linker peptide (Lord, 1985a; Lamb et al, 1985). This ricin linker is similar to the linker separating the 2S albumin subunits, not in terms of length, but both have a serine residue at the amino terminus and an asparagine residue at the carboxyl terminus. It could be that one or more common specific endoproteases are responsible for the removal of both linkers.

It would seem that for many processing events involving post-translational modification of plant storage proteins that either a short linker peptide is removed (Higgins et al, 1983; Bowles et al, 1986) or a single cleavage is made of a peptide bond (Higgins, 1984; Boulter, 1984). In both instances, cleavage on the carboxyl side of an asparagine residue is a widespread feature (Lord and Robinson, 1986). This would explain cleavage to liberate the large subunit of the 2S albumin and the small subunit from the remaining polypeptide. The linker would still need to be processed away from the small subunit by cleavage on the amino terminal side of SER 170. In determining the cleavage sites of the Arabidopsis thaliana 2S albumin, Krebbers et al (1988) noted that proximal to the small subunit of both the Arabidopsis 2S albumin and the Brassica napin was conserved THR-ASN. This is seen preceding the small subunit sequence in the preproalbumin here (residues 134-135). They also noted that cleavage between the carboxyl terminus of the small subunit and the linker peptide followed no discernible pattern. The end of the linker was cleaved with a consensus of GLU-ASN followed by PRO-GLU in the large subunit. This is not conserved in the preproalbumin here, although an asparagine marks the end of the linker

peptide, and a proline marks the beginning of the small subunit. Unlike the Arabidopsis albumin, no processing is required of the carboxyl terminus of the preproalbumin since this also defines the end of the large subunit. Processing of the carboxyl terminus of the large subunit takes place in the Arabidopsis albumin and the Brassica napin with serine conserved as the last residue of the large subunit.

The linker between the two subunits of the 2S albumin is small in comparison to those in other 2S albumins and ricin: the napin of Brassica napus has a 19 residue linker (Crouch et al, 1983), which shows homology to the linker of the 2S albumin from Arabidopsis thaliana, following a haaahahaaahaN motif (h, hydrophobic; a, acidic; N, asparagine: Krebbers et al, 1988). The Brazil nut sulphur-rich preproalbumin has a linker of only five residues though, more in line with that observed with the preproalbumin here (Altenbach et al, 1987).

Polyadenylation signals

In animal genes the consensus sequence for the polyadenylation signal is AATAAA (Proudfoot and Brownlee, 1976), which appears 15-30 bases away from polyadenylation sites and is important not only in directing the addition of a poly A tail to the end of a transcript but also to dictate the point of transcription termination (Whitelaw and Proudfoot, 1986). Plants however appear not to be so conserved in their use of this consensus signal (Lycett et al, 1983; Heidecker and Messing, 1986). Legumin cDNAs show an overlapping repeat sequence AATAAATAAA 19 bases from the poly A site. Leghaemoglobin C also contains this sequence (Hyldig-Nielsen et al, 1982), as does zein as AATAATAAA (Geraghty et al, 1981) but these are not present in the expected position but 104 and 56

bases away respectively. Variants of the normal sequence GATAAA and AATAAGAAA are found in the expected positions, suggesting perhaps that these act as plant specific signals. The cDNA of thaumatin (Lazaro et al, 1988), whose amino acid sequence was reported to have homology with the castor bean 2S albumin (Sharief and Li, 1982) has three repeats of the signal AATAAA. It has been said that this type of multiple naturally occurring polyadenylation signal is connected with variability in the site of polyadenylation (Tosi et al, 1981).

In Xenopus, there are two forms of the beta-1-globin mRNA that differ in their site of polyadenylation. A minor site is located 46 nucleotides downstream of the major site and is used in 21% of mRNA molecules. If the major site is deleted then the minor site is used in over 90% of mRNA molecules. If the major site is changed to AATACA, 35% of transcripts continued to be polyadenylated from the major site, suggesting a high degree of flexibility in the precise polyadenylation signal (Mason et al, 1985).

The preproalbumin cDNA has two consensus signals AATAAA positioned at bases 816 and 1076. There is also a near consensus signal AATAAG at base 922. Following the polyadenylation signal at base 816 is a hairpin loop structure (Figure 17). These structures have been noted in zein and Brassica napin genes (Heidecker and Messing, 1986; Josefsson et al, 1987) and their role, if any, in defining the 3' end of mRNAs is unclear (Birnstiel et al, 1985) although they may be involved in termination of specific sets of genes (Hentschel and Birnstiel, 1981). Clone 14g4 was found to be polyadenylated 21 residues from the consensus sequence at base 1076, whilst clone 8g8 was polyadenylated 139 bases away from the

consensus sequence at base 816. However this was only 26 bases away from the near consensus sequence AATAAG at base 929. It is not certain which of the two signals operated in the polyadenylation of clone 8g8. Site directed mutagenesis and microinjection of *in vitro* generated transcripts into *Xenopus* oocytes may reveal the answer. What seems likely is that the polyadenylation signal at position 1076 is an inefficient one. Since most of the cDNAs cloned were of group 1 category, that is shorter than clone 14g4 and missing the third polyadenylation signal (base 1076), one of the first two signals at positions 816 and 922 must be used preferentially. It could be of course that different gene members generated transcripts of different lengths and one or more genes produced transcripts of category 1 at a greater rate than genes producing transcripts of category 2. Only a full characterisation of all cDNA members and the corresponding genes will enable this to be clarified further.

Figure 19

Predicted secondary structure for the polypeptide derived from the open reading frame sequence of clone 14g4

The amino acid sequence of the open reading frame generated from the nucleotide sequence of clone 14g4 was subject to the secondary structure prediction algorithm of Chou and Fasman (1978ab).

The amino terminus is indicated by NH_2, whilst the carboxyl terminus is indicated by COOH. Numbering is from the first residue of the open reading frame. Potential glycosylation sites are indicated by pinned circles.

alpha helix

beta sheet

turns

coils

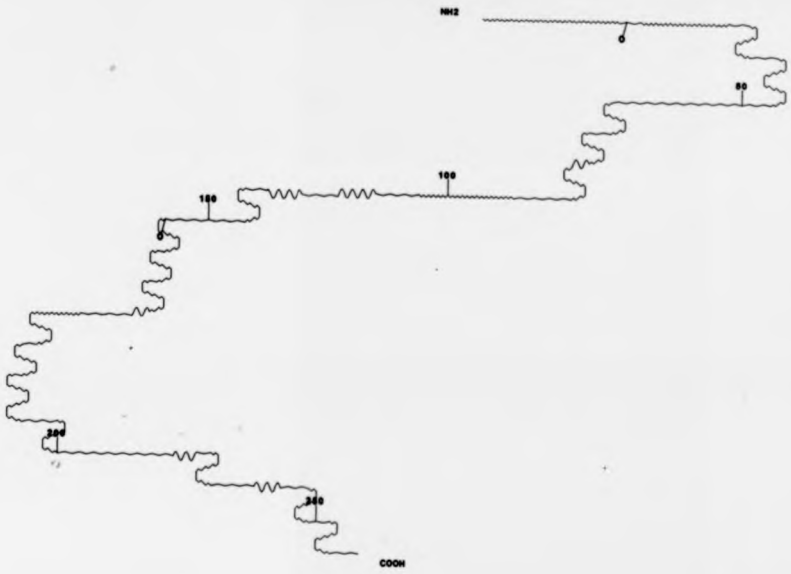

NH2
50
100
150
200
250
COOH

Figure 20

Hydrophobicity plot of the open reading frame polypeptide derived from the nucleotide sequence of clone 14g4

The figure displays the distribution of polar and apolar residues along the sequence of the polypeptide derived from clone 14g4. The polypeptide sequence is represented by its predicted secondary structure (Figure 19). Hydrophobic residues are represented by diamonds whilst hydrophilic residues are represented by octagons. The size of the symbols is proportional to the value of the attribute. The plot was generated from the algorithm of Kyte and Doolittle (1982), using a seven residue window.

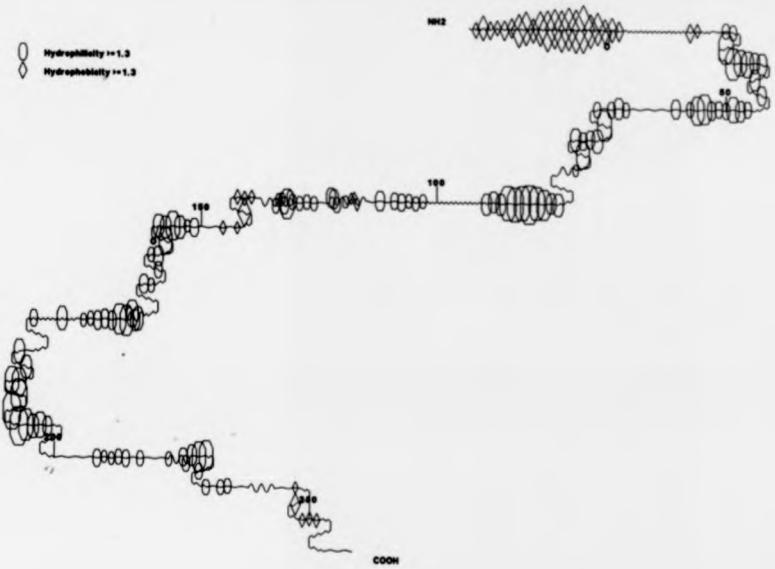
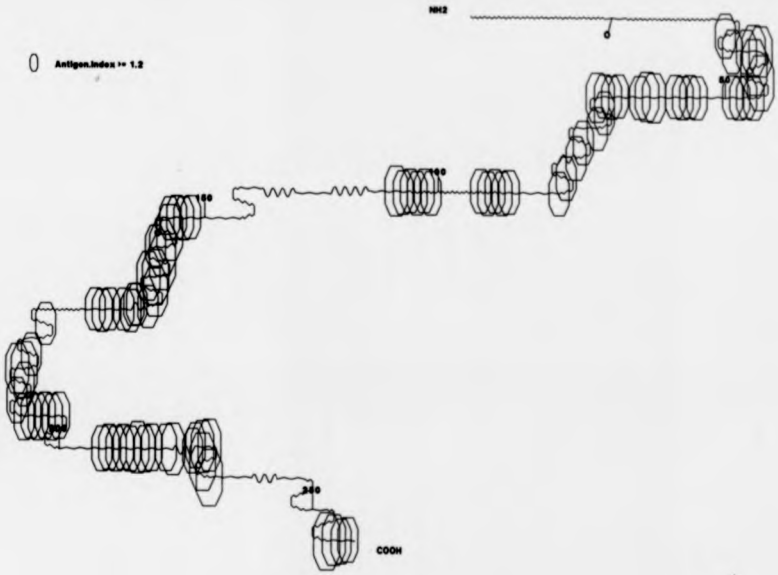

Hydrophilicity >= 1.3
Hydrophobicity >= 1.3
NH2
0
50
100
150
200
240
COOH

Figure 21

Antigen index predicted for the open reading frame polypeptide derived from the nucleotide sequence of clone 14g4.

The figure shows the predicted distribution of antigenic sites along the sequence of the polypeptide derived from clone 14g4. The polypeptide sequence is represented by its predicted secondary structure (Figure 19). The size of the octagons is proportional to the value of the attribute. The prediction was generated from the algorithm of Jameson and Wolf (1988).

Facing -192-


Antigen.Index >= 1.2
NH2
0
60
120
180
240
300
340
COOH

D2.4 Predictions on the structure of the preproalbumin

The open reading frame was subject to algorithms which attempted to define the secondary structure (Figure 19), hydrophilicity (Figure 20) and antigenicity (Figure 21) of the preproalbumin.

Secondary structure

The Chou-Fasman method was used to produce the secondary structure prediction. In essence this analyses the residues of the polypeptide noting to which class each residue belongs. Rules determine the structure adopted, whether it be alpha-helix, beta-sheet, turn or coil. No attempt will be made to explain the algorithm since it has been defined by its authors (Chou and Fasman, 1978ab). What must be said however, is that any secondary structure prediction must be treated with caution, and no attempt must be made to envisage it as a tertiary structure. This is especially true for the preproalbumin. It has already been noted that the polypeptide has anomalous migratory properties on polyacrylamide gels, probably due to its high intrachain disulphide bonding (D2.3). Certainly it is known that the 2S albumin is composed of a large subunit and a small subunit linked by one or possibly two disulphide bonds. This would account for the two cysteine residues located in the small subunit and two of the six residues located in the large subunit. The structures of the Bowman-Birk type serine protease inhibitor from soybean and the sulphur rich seed protein of Lupinus angustifolius L. conglutin-delta have been determined along with the positions of the disulphide bridges (Odani and Ikenaka, 1973; Lilley and

Inglis, 1986). Although the mature 2S albumin is a serine protease inhibitor (McGurl, 1986) it shows greater homology to conglutin-delta especially in terms of its cysteine residue positions and subunit composition. Conglutin-delta possesses a structure strong in alpha-helix (38%). Both cysteine residues of the small subunit are involved in cross-linking to the large subunit whilst the remaining cysteine residues of the large subunit are involved in cross-linking amongst themselves to stabilise the helical structure. There is a free sulphydryl group associated with conglutin-delta which is probably not present in the castor bean 2S albumin since the latter has one less cysteine residue.

It would seem that the secondary structure prediction derived here (Figure 19) should be treated with some scepticism, since it is only a representation of the sequence of residues obeying a simple algorithm. It pays no attention to the location and pairing of cysteine residues and does not draw upon the structures of proteins that have already been determined such as conglutin-delta. Apart from the prediction of the structure of the mature 2S albumin, the rest of the prepropeptide must be treated with similar caution, since it should be noted (and will be discussed shortly) that it contains a further eight cysteine residues in positions similar to those found in the mature 2S albumin.

As for the positions of the glycosylation sites predicted at sequences ASN-X-SER/THR (Pless and Lennarz, 1977; Hart *et al*, 1979), it should be noted that a glycosylation site appearing in the signal peptide seems highly improbable since the latter is cleaved whilst emerging into the lumen of the endoplasmic reticulum before glycosylation takes place. The

second predicted site occurs at the asparagine which borders the cleavage site of the small subunit of the mature 2S albumin. Whether glycosylation takes place, and if so whether this is part of a signal to cause cleavage of the mature 2S albumin in the protein bodies, is unknown. All that may be said of the prediction is that it was correct in determining no glycosylation sites in the mature 2S albumin (Tully and Beevers, 1976; McGurl, 1986).

Hydrophilicity plot

The hydrophilicity plot predicted for the preproalbumin (Figure 20) shows a number of interesting points. Firstly it correctly shows the hydrophobic nature of the signal sequence. Since the functions of the signal are to anchor the ribosome to the cytosolic face of the endoplasmic reticulum and deposit the forming polypeptide into the lumenal side of the endoplasmic reticulum, it needs to span the membrane and to do so requires a high hydrophobic content. This is indeed seen, with the hydrophobicity of the preproalbumin being lost shortly after the predicted site of signal cleavage.

Secondly, the remainder of the preproalbumin (the proalbumin) appears very hydrophilic, and this is no doubt a reflection of the high nitrogen containing amino acids such as glutamine contained within the polypeptide. The hydrophilicity is also reflected in the mature protein's high solubility profile and its ability to dissolve in water alone.

Thirdly, there are two short areas of hydrophobicity contained within the proalbumin. One area is located at the carboxyl end of the 2S

albumin large subunit sequence whilst the other is located in the unassigned sequence immediately prior to the small subunit sequence. If the structure of the mature 2S albumin is like that of conglutin-delta then the hydrophobic sequence would be positioned at the centre of the alpha-helix, shielded from the aqueous phase. This hydrophobic sequence could be important then in the initial conformational arrangement, ensuring cystaine residues are correctly aligned for disulphide bridging.

The hydrophilicity plot was generated using a seven residue window with the algorithm of Kyte and Doolittle (1982). This is generally used for finding surface exposed regions that may be antigenic sites. A longer window of 19-21 residues would have shown longer hydrophobic membrane spanning segments more clearly, but was not appropriate for such a small hydrophilic polypeptide.

Antigenicity

The antigen index predicted for the preproalbumin (Figure 21) is an extension of the hydrophilicity plot, areas of high hydrophilicity being the most usual areas recognised by antibody (Hopp and Woods, 1981). This suggests as expected from such a hydrophilic protein that large parts of the proalbumin are capable of inducing an antigenic response. This is credible since it is known that the mature 2S albumin is one of the most potent allergenic substances known (Youle and Huang, 1978b; McGurl, 1986). McGurl also noted that the majority of the antigenicity associated with the mature 2S albumin was conformation dependent, reflecting the point that the major epitopes of the protein are probably discontinuous. Hence, whilst the antigen index serves to suggest that

the protein has a large number of stretches of residues capable of forming epitopes, these are in fact only continuous epitopes and so do not truly reflect the nature of the mature protein.

Figure 22

A putative second storage protein occurs on the open reading frame polypeptide derived from the nucleotide sequence of clone 14g4.

The nucleotide and derived amino acid sequence of clone 14g4 is shown as before (Figure 15). The small and large polypeptide sequences of the 2S albumin purified and sequenced by Sharief and Li (1982) are boxed. Additional putative, glutamine-rich 2S albumin small and large subunits are underlined and coloured green and yellow respectively. Sequences which may be removed by post-translational processing are underlined with a broken line. Cysteine residues are highlighted by a second underline. An arrow defines the putative signal peptidase cleavage site, from which numbering begins.

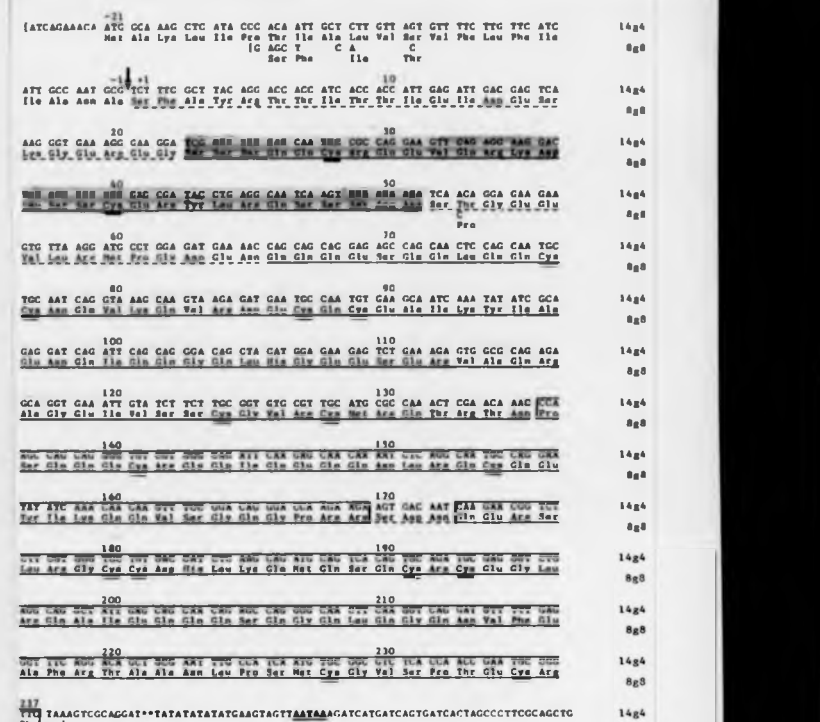
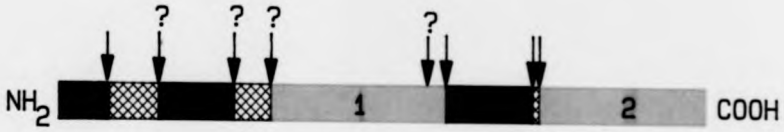


Figure 23

Schematic representation of the proposed regions of the preproprotein containing the previously identified 2S albumin.

The preproprotein sequence derived from the nucleotide sequence of clone 14g4 is subdivided into peptide moities. The sites of post-translational processing are indicated by arrows, with question marks associated with those process steps not positively identified. The amino terminus is represented by NH_2 whilst the carboxyl terminus is represented by COOH. The subunits labelled '2' are those of the previously purified and sequenced castor bean 2S albumin (Sharief and Li, 1982). Those labelled '1' are those of the proposed second storage protein.

The figure is drawn to scale where 1 cm represents 14 residues.



SIGNAL PEPTIDE

SMALL SUBUNIT

LINKER PEPTIDE

LARGE SUBUNIT

1: PUTATIVE STORAGE PROTEIN

2: IDENTIFIED STORAGE PROTEIN

PROCESSING SITE

D2.5 Examination of the preproalbumin area as yet unaccounted for

So far the regions of the preproalbumin accounted for include: the signal sequence between residues -21 to -1; the small subunit of the previously characterised 2S albumin, between residues 136 and 169; the large subunit from residues 173 to 237; and a small linker peptide removed from between the subunits, residues 170 to 172. This accounts for 123 of the 258 residues of the polypeptide (47.7%). This means that over half of the polypeptide has so far been unaccounted for. Large precursors for 2S albumins are not unusual, and most are usually post-translationally processed away to generate the mature protein. As mentioned (I3.2) the precursor of Brassica napin (which is composed of a 9 kDa and 4 kDa subunit) is 20 kDa, that for the Brazil nut methionine rich albumin (9 kDa and 3 kDa) is 20 kDa and that for the sunflower seed albumin (19 kDa monomer) is 38 kDa. In the case of the castor bean, processing appears to be particularly wasteful if a precursor of actual molecular weight 29 kDa generates only a heterodimer consisting of a 4 kDa and 7 kDa subunit (Crouch et al. 1983; Ericson et al. 1986; Altenbach et al. 1987; Allen et al. 1987).

Closer examination of the unidentified sequence reveals that a large number of residues are glutamine and there also appears to be a richness of cysteine residues. This indeed could mean that the 135 residues within the preproalbumin not accounted for represent another 2S albumin storage protein (Figure 22). This would make sense. Seed storage proteins, especially the 2S albumins in some plants, are deposited so

that the germinating seedling has sources of nitrogen and sulphur. It would make sense then for the plant to be efficient in this deposition process during seed development. If two storage proteins can be sent to the protein bodies for deposition using only one precursor molecule, then only one set of signal sequences and linkers need be processed and wasted. As mentioned (I3.2) no other 2S albumin precursor so far analysed has been large enough to accommodate a second storage protein, except for the sunflower albumin. It was mentioned by Allen et al (1987) that possibly a second storage protein could exist on the sunflower albumin precursor, but this was not elaborated further. Also the sunflower albumin is different from most 2S albumins in that the mature protein (and presumably also the putative second storage protein) is a single chain polypeptide.

If a second storage protein is present on the castor bean preproalbumin then it must be processed to generate a small and large subunit. In defining the processing sites of the mature protein it was already discussed how consensus sequences are difficult to obtain (D2.3). What is known is that in the castor bean, linker polypeptides between subunits appear to be delineated by an amino terminal serine residue and a carboxyl terminal asparagine residue. Asparagine is also implicated in the cleavage of the small subunit at the amino terminus. Analysis of the sequence shows that the large subunit could be defined between residues 66 and 135. This would ensure that all six cysteines are present and that cleavage of the small subunit of the previously characterised 2S albumin would liberate the carboxyl terminus (although this does not prevent another processing event trimming the carboxyl terminus as long

as the cysteine at position 128 remained with the large subunit). The amino terminus would be cleaved in a manner similar to the cleavage in the previously characterised 2S albumin, on the carboxyl side of asparagine, with glutamine as the first residue of the subunit.

The linker region could be defined between residues 52-65, producing a 14 residue peptide bounded by serine and asparagine. This would define the carboxyl terminus of the small subunit as ARG-ARG, already seen in the previously characterised 2S albumin.

The site of cleavage at the amino terminus of the putative small subunit is not so well defined but has been tentatively assigned to residue SER-23 (or SER-24/25). No asparagine is present between the site of signal cleavage and CYS-28 which should be present on the small subunit. A linker region is usually removed between the amino termini of 2S proalbumins (that is, after signal cleavage) and the small subunit and this has therefore been tentatively assigned between residues 1 and 22. A simplified schematic diagram of the preproalbumin and its known and unknown processing sites is given in Figure 23.

Table 2

Quantity of glutamine and cysteine residues in the proposed regions of the preproprotein

The proposed regions of the preproprotein are tabulated and the number of glutamine and cysteine residues contained within each are given.

Region numbers refer to amino acid residues from the point of predicted signal peptidase cleavage.

An identity to each region is given where possible: 'SS1' and 'LS1' refer to the small and large subunits of the putative storage protein, whilst 'SS2' and 'LS2' refer to the subunits of the previously identified storage protein. 'Linker' refers to the region between a small and large subunit.

REGION	IDENTITY	RESIDUE	
		GLUTAMINE	CYSTEINE
-21 - -1	SIGNAL	0	0
1 - 22	?	0	0
23 - 51	SS1	5	2
52 - 65	LINKER	0	0
66 - 135	LS1	16	6
136 - 169	SS2	11	2
170 - 172	LINKER	0	0
173 - 237	LS2	12	6

Figure 24

Sequence comparison of the castor bean 2S albumin small subunits with other plant storage proteins

A comparison is made between a region of the putative castor bean small subunit and similar regions in other plant storage proteins.

1 Putative castor bean small subunit, residues 23-51

2 Identified castor bean small subunit, residues 1-30 (Sharief and Li, 1982)

3 Brazil nut 2S sulphur-rich albumin, residues 35-64 (Altenbach et al. 1987)

4 Napin small subunit from Brassica napus. residues 41-71 (Crouch et al. 1983)

5 Pea seed albumin, residues 1-33 (Higgins et al. 1986)

6 Wheat alpha-amylase inhibitor, residues 1-33 (Kashlan and Richardson, 1981)

7 Wheat albumin, residues 1-29 (Shewry et al. 1984)

8 Millet trypsin inhibitor, residues 1-29 (Campos and Richardson, 1983)

Sequences were aligned to maximise homology. Coloured boxes indicate homology found in plant proteins against the putative castor bean albumin only. Standard one letter code for amino acids is used. Residue numbers used are the respective authors' originals. Data modified from Kreis et al (1985b) and Higgins et al (1986).

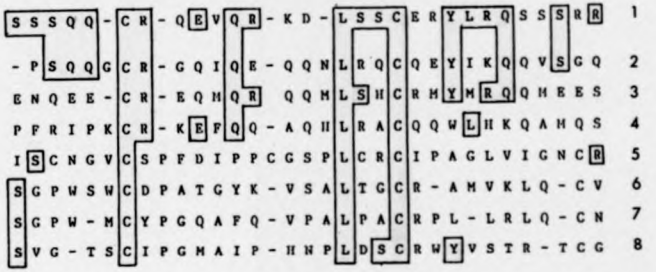

SSSQQ-CR-QEVQR-KD-LSSCERYLRQSSSRR 1
-PSQQGCR-GQIQE-QQNLRQCQEYIKQQVSGQ 2
ENQEE-CR-EQHQR QQHLSHCRHYMRQQHEES 3
PFRIPKCR-KEFQQ-AQHLRACQQWLHKQAHQS 4
ISCNGVCSPFDIPPCGSPLCRCIPAGLVIGNCR 5
SGPWSWCDPATGYK-VSALTGCR-AMVKLQ-CV 6
SGPW-MCYPGQAFQ-VPALPACRPL-LRLQ-CN 7
SVG-TSCIPGMAIP-HNPLDSCRWYVSTR-TCG 8

Figure 25

Sequence comparison of the castor bean 2S albumin large subunits with other plant storage proteins

A comparison is made between a region of the putative castor bean large subunit and similar regions in other plant storage proteins.

1 Putative castor bean large subunit, residues 73-108

2 Identified castor bean large subunit, residues 177-212 (from clone 14g4 open reading frame)

3 Sunflower albumin consensus (Allen et al. 1987)

4 Napin large subunit from Brassica napus, residues 102-136 (Crouch et al. 1983)

5 Barley trypsin inhibitor, residues 40-75 (Odani et al. 1983)

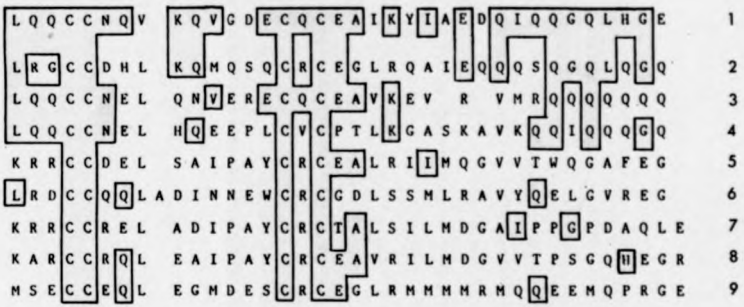
6 Wheat a-amylase inhibitor, residues 39-75 (Maeda et al. 1983)

7 Maize trypsin inhibitor, residues 43-80 (Mahoney et al. 1984)

8 Millet trypsin inhibitor, residues 41-77 (Campos and Richardson, 1983)

9 Brazil nut 2S sulphur-rich albumin, residues 78-114 (Altenbach et al. 1987)

Sequences were aligned to maximise homology. Coloured boxes indicate homology found in plant proteins against the putative castor bean albumin only. Standard one letter code for amino acids is used. Residue numbers used are the respective authors' originals. Data modified from Kreis et al (1985) and Allen et al (1987).


LQQCCNQV KQVGDECQCEAIKYIAEDQIQQGQLHGE 1
LRGCCDHL KQMQSQCRCEGLRQAIEQQQSQGQLQGQ 2
LQQCCNEL QNVERECQCEAVKEV R VMRQQQQQQQQ 3
LQQCCNEL HQEEPLCVCPTLKGASKAVKQQIQQQGQ 4
KRRCCDEL SAIPAYCRCEALRIIMQGVVTWQGAFEG 5
LRDCCQQLADINNEWCRCGDLSSMLRAVYQELGVREG 6
KRRCCRELADIPAYCRCTALSILMDGAIPPGPDAQLE 7
KARCCRQLEAIPAYCRCEAVRILMDGVVTPSGQHEGR 8
MSECCEQLEGMDESCRCEGLRMMMMRMQQEEMQPRGE 9

D2.6 Evidence for the existence of a second storage protein from sequence data

The sequence data generated so far from cDNA clone 14g4 suggests that a second storage protein may be present on the preproalbumin that also possesses the previously characterised 2S albumin, both being directed to the endoplasmic reticulum as one preproalbumin by a cleavable signal sequence, from which the proalbumin is directed to the protein bodies with subsequent endoproteolytic cleavage at specific sites to generate two different heterodimeric 2S albumin storage proteins (Lord 1985a and subsection D2.5). To avoid confusion the previously characterised 2S albumin shall be referred to as such or as storage protein 2. The putative storage protein shall be known as such or as storage protein 1. They are numbered as to the order in which they are present from the amino terminus of the preproalbumin (Figure 23). Reference to the subunits shall be done by abbreviating small subunit and large subunit to SS and LS respectively followed by the number of the storage protein concerned. For example, the small subunit of the putative 2S albumin shall be known as SS1.

Using the speculative assignment of processing sites, the sizes of the proteins generated would be 11,306 Da for the previously characterised 2S albumin, and 11,530 Da for the putative storage protein. The subunit sizes are similar also: LS1, 8041 Da; LS2, 7349 Da; SS1, 3489 Da; SS2, 3957 Da.

The previously characterised 2S albumin has a total of 8 cysteine residues and 23 glutamine residues, whilst the putative 2S albumin has a total of 8 cysteine residues and 21 glutamine residues. The portions of the precursor removed during processing (a total of 60 amino acids derived from the signal sequence, the post-translationally removed N-terminal peptide, and the linker peptides) contain no cysteine or glutamine residues (Table 2). The distribution of cysteine residues is also conserved with two present in the small subunits and six in the large subunits. This arrangement is well conserved in many other 2S albumins and related cereal prolamins (Kreis et al, 1985b).

The function of the post-translationally removed N-terminal peptide is unknown, but it may contain a protein body targeting sequence. Post-translational N-terminal processing has been implicated during the biosynthesis of other plant protein body/vacuolar proteins (Crouch et al, 1983; Graham et al, 1985; Hattori et al, 1987).

Homologies between the putative small and large subunits have been made with other 2S albumins, trypsin inhibitors and storage proteins in the areas defined by Kreis et al (1985b), as the most conserved regions (region A in the small subunit and region B in the large subunit). These regions of sequence conservation are centred around the arrangement of cysteine residues, suggesting a structural homology, rather than a strict sequence homology (Kreis et al, 1985ab; Krebbers et al, 1988; Allen et al, 1987; Altenbach et al, 1987; Shewry et al, 1984).

The putative small subunit shares homology with the previously characterised castor bean 2S albumin small subunit and other seed

storage proteins known to be part of a superfamily. Figure 24 shows a limited sequence comparison in region A for these polypeptides and highlights residues in SS1 that are found in other members of the group, including the two cysteine residues (CYS-28 and CYS-40) and the single leucine residue (LEU-37) that are an invariant feature of these peptides.

The putative large subunit, LS1, likewise shares homology with the previously characterised 2S albumin large subunit and related polypeptides within region B (Figure 25). In this case the most striking feature is the CYS-CYS-(X_9)-CYS-(X)-CYS motif (residues 76-89 and 180-193 in the preproalbumin).

It was also suggested earlier that the predicted secondary structure showed similarities between the two parts of the preproalbumin. In particular, both storage protein 1 and 2 shared high hydrophilicity profiles except at the carboxyl terminus of their large subunits. Like the large subunit of the previously characterised 2S albumin, the carboxyl terminus of the large subunit of the putative 2S albumin may be involved in generating the final three dimensional structure of its protein.

Early studies on the castor bean protein body proteins suggested that the 2S albumins were composed of two major proteins of apparent molecular weights 10.5 and 12 kDa (Youle and Huang, 1978a; Tully and Beevers, 1976; Gifford et al. 1982). Only one of these was characterised and sequenced (Li et al. 1977; Sharief and Li, 1982). It is possible that the other 2S albumin present in the protein bodies of castor bean

seeds is the same as the putative 2S albumin predicted for the preproalbumin. Direct N-terminal sequencing of this purified 2S albumin would provide positive identification and define the sites of proteolytic cleavage from the precursor.

D2.7 Conclusions

The castor bean preproalbumin has been cloned and sequenced. The mature 2S albumin is positioned at the carboxyl end of the preproalbumin, and the sequence corrects the protein sequence published previously. Located also in the preproalbumin is a signal sequence, and a further glutamine and cysteine rich area which could represent the second 2S albumin present in the protein bodies of the castor bean, as yet undefined.

D3 EVIDENCE FOR A SECOND 2S ALBUMIN IN CASTOR BEANS

D3.1 Summary of results

The preproalbumin and *Brassica napus* prepronapin polypeptides generated from a wheatgerm *in vitro* translation system did not cross immunoprecipitate using rabbit polyclonal antibodies raised against the mature proteins. Located within the 2S albumin fraction from protein bodies of the castor bean were 5 major polypeptides, 3 of which were attributed to the 2S albumin previously characterised by Sharief and Li (1982). This 2S albumin was found as a doublet of proteins, composed of a large subunit disulphide linked to one of two differently sized small subunits. Rabbit polyclonal antibodies raised towards the characterised 2S albumin did not recognise the remaining 2 polypeptides. *In vitro* translation products of transcripts generated *in vitro* from preproalbumin cDNA clone 14g4 were processed by elements contained within a castor bean protein body extract, but not fully to the correct subunit sizes.

Figure 26

Cross-immunoprecipitation study of castor bean preproalbumin and Brassica napus prepronapin.

Transcripts of the castor bean preproalbumin and the prepronapin of Brassica napus were generated by in vitro transcription of cDNA clones 14g4 and pN2 subcloned into pGEM 3blue (M3.1.1 and figure 10, except the pN2 clone was subcloned into the Pst I site). These were translated in vitro using a wheatgerm system (M2.2.b). 8 ul aliquots of the reaction products were subject to immunoprecipitation with rabbit polyclonal antibody raised towards either the Brassica napin or the castor bean albumin. Products were resolved on a 15% (w/v) denaturing polyacrylamide gel under reducing conditions (M1.1.a), after which the gel was fluorographed, dried and exposed to film overnight at -70°C (M2.2.a).

Lane 1 — 8 ul preproalbumin translation products

Lane 2 — 8 ul preproalbumin immunoprecipitated with 3 ul antiserum raised towards the 2S albumin

Lane 3 — 8 ul preproalbumin immunoprecipitated with 3 ul antiserum raised towards the napin

Lane 4 — 8 ul prepronapin translation products

Lane 5 — 8 ul prepronapin immunoprecipitated by 3 ul antiserum raised towards the napin

Lane 6 — 8 ul prepronapin immunoprecipitated by 3 ul antiserum raised towards the castor bean preproalbumin

Antibodies directed towards the napin were a kind gift from Dr. A. Ryan. The napin precursor cDNA pN2 was a generous gift from Dr. M. Crouch as described in Crouch et al (1983).


1 2 3 4 5 6
kDa
180
116
84
58
48.5
36.5

D3.2 Antigenic crossreactivity between 2S albumins

Figure 26 shows the result of an experiment that was aimed at determining whether the preproalbumin of castor bean and the prepronapin of *Brassica napus* were antigenically related. By translating *in vitro* transcripts generated from cDNA clones in a wheatgerm *in vitro* system, polypeptide products were immunoprecipitated with rabbit polyclonal antibodies raised either towards the mature characterised 2S albumin of castor bean or the mature *Brassica* napin. The result seems to suggest that the two precursors show no common antigenic determinants, at least not by the sensitivity of this experiment.

This may seem surprising since the mature proteins (both putative and characterised 2S albumins from castor bean, and the napin) are known to possess homologies in their primary structures (Figures 23 and 24). Nevertheless, as has been pointed out, the homologies in the conserved regions (Kreis *et al*. 1985b) are known to be related at the positions of the cysteine residues thereby ensuring an overall structural similarity between 2S albumins and related storage proteins. What is clear when comparing 2S albumin sequences is that the majority of a sequence is not homologous (40% homology is significant), suggesting that the majority of epitopes should not be shared.

The experiment shows the *in vitro* generated polypeptides of the two precursors are not antigenically related. However the folding of the polypeptides in an *in vitro* system in the presence of the reducing agent dithiothreitol, thus preventing disulphide bonding, is likely to

generate structures dissimilar to the prepropeptides or mature proteins generated in the plant. Hence, whilst this experiment shows no shared determinants, it is still quite possible that the mature proteins may be crossreactive.

Any further studies into the crossreactivity of the preproalbumins should be performed by *in vivo* labelling of developing seeds followed by immunoprecipitation, to ensure correctly folded structures, and by analysis of the mature proteins by Western blotting or enzyme-linked immunosorbent assays.

It should also be noted that antibodies raised towards the *Brassica* napin in failing to recognise the preproalbumin, must have found no common antigenic determinants on the putative storage protein located in the preproalbumin. Since the primary structure homologies are no greater between the characterised 2S albumin and the putative 2S albumin, than with the *Brassica* napin it might well be that no common antigenic determinants are shared between the two albumins (characterised and putative) from the castor bean plant.

Few studies have been performed on the crossreactivity of the 2S albumins. Templeman *et al* (1986) showed ostrich fern albumin storage proteins shared antigenic determinants and nucleotide sequence homology with *Brassica* napin. Ferns diverged from the evolutionary line giving rise to angiosperms prior to the divergence of monocotyledonous and dicotyledonous plants. More extensive studies have been performed on other storage protein groups to investigate antigenic crossreactivity. Dudman and Millerd (1975) noted that proteins related to the legumin of

Vicia faba were widely related amongst the taxa of Fabeae and Trifolieae, whilst the vicilin was immunochemically different to most genera tested (except to Pisum sativum). Doyle et al. (1985) showed the 7S storage proteins of soybean were related to globulin enriched proteins within the tribes of Phaseoleae and Papilionoideae suggesting evolutionary relatedness as well as strong selectional restraint. Robert et al (1985) showed homology amongst 12S legumin-like polypeptides in cereals and pea. Guldager (1978) showed vicilin and legumin of Pisum sativum were immunologically distinct. Dierks-Ventling and Ventling (1982) showed common antigenic sites existed for the prolamins of wheat, barley and maize whilst Festenstein et al (1984) showed that antigenic determinants in most prolamins were shared in the sub-family Festucoideae (which includes barley) of the Gramineae, whilst the prolamins of the sub family Panicoideae were shown not to be homologous with the former. It is thought that the conserved highly repetitive domains present in the prolamins yet absent in the 2S albumins are responsible for the crossreactivity.

The most extensive study of its kind was carried out by Konarev et al (1987) who studied the seed storage proteins of 2000 cultivars from 300 species of 70 genera by immunoprecipitation methods, in order to generate data on protein evolution. Their results showed that plant storage proteins were more changeable in evolution than other plant proteins such as enzymes and histones. They established that storage proteins from the seeds of one species were identical as antigens, those from species and genera within families possessed partial identity,

whilst those from different families were non-identical. The 2S albumins were not studied.

Although the literature appears contradictory at times, it would seem that antigenic determinants exist between similarly grouped storage proteins found in plants that are both closely and distantly related. As for the 2S albumins it might be that shared antigenic determinants exist between those whose primary sequences are more closely related.

Figure 27

Water-soluble components of the castor bean protein bodies

The water-soluble components of the castor bean protein bodies were extracted and the 2S albumin fraction separated as described (M1.4). The previously characterised 2S albumin (Sharief and Li, 1982) was purified as described (M1.3). Samples of these proteins were resolved on a 16.5% (w/v) denaturing polyacrylamide gel buffered with tricine (M1.1.b) under reducing conditions. The gel was stained in Coomassie brilliant blue R as described (M1.2.a).

Lane A 10 ug previously identified 2S albumin

Lane B 30 ul 2S albumins, post G-50 Sephadex column

Lane C 7 ul total soluble proteins of the protein bodies

The position of the 7S lectins is marked '7S'. 'LS' and 'SS' are abbreviations of large subunit and small subunit. Marker proteins were as described (M1.1.b).

Facing -217-

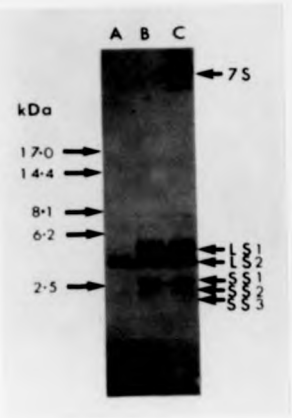

A B C
7S
kDa
17·0
14·4
8·1
6·2
2·5
LS1
LS2
SS1
SS2
SS3

Figure 28

Western blot of the subunits of the previously characterised 2S albumin

The total soluble proteins of the castor bean protein bodies and the previously characterised 2S albumin (Sharief and Li, 1982) were resolved on a 16.5% (w/v) denaturing polyacrylamide gel buffered with tricine (M1.1.b) under reducing conditions. The proteins were transferred to nitrocellulose by Western blotting and probed with rabbit antiserum containing polyclonal antibody directed to the characterised 2S albumin as described (M1.6).

Lane 1 7 ul total soluble protein body proteins

Lane 2 10 ug purified 2S albumin

'LS' and 'SS' refer to large subunit and small subunit. Prestained molecular weight markers were resolved on the gel and their subsequent presence on the nitrocellulose indicated correct transfer. Their molecular weights though were unsuitable for this resolution and are not indicated.

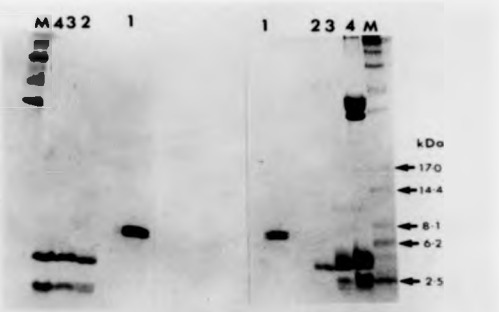

1 2
LS2
SS2
SS3

Figure 29

Western blot of the total soluble protein body proteins.

Total soluble protein body proteins, separated 2S albumins (M1.4) and purified 2S albumin (M1.3) were resolved on two identical 16.5% (w/v) denaturing polyacrylamide gels buffered with tricine under reducing conditions. The purified albumin was run under non-reducing conditions also. One gel was blotted onto nitrocellulose whilst the other was stained in Coomassie brilliant blue R (M1.2.a). The nitrocellulose was probed with rabbit antiserum containing polyclonal antibody raised towards the previously characterised 2S albumin (Sharief and Li, 1982 as described M1.6).

Lane 1	10 ug purified 2S albumin (1 ug run on the gel to be blotted) under non-reducing conditions
Lane 2	10 ug purified 2S albumin run under reducing conditions
Lane 3	30 ul 2S albumins run under reducing conditions
Lane 4	7 ul total soluble protein body proteins run under reducing conditions
Lane M	Prestained and low molecular weight markers.



WESTERN BLOT COOMASSIE STAINED GEL

D3.3 Observations on the castor bean seed protein body matrix proteins

The soluble matrix of the castor bean protein bodies was analysed using denaturing polyacrylamide gels buffered with tricine under reducing conditions. When tricine was used instead of glycine the resolving power was greatly enhanced for very small proteins and peptides from sizes of 1 kDa upwards (Shagger and von Jagow, 1987). The only drawback to using this system was the consistently spurious running of molecular weight markers in relation to the 2S albumins. This was probably more of a reflection of the high cysteine content of the 2S albumins (Masumura et al, 1989). Figure 27 shows the resolution of the proteins from the total soluble matrix (Figure 27, lane C), the 2S albumins purified from the total soluble matrix by gel filtration (Figure 27, lane B) and the previously characterised 2S albumin (Figure 27, lane A). The total soluble matrix proteins can be seen to include the 7S lectins as well as the 2S albumins. The 7S lectin fraction contains the subunits of ricin and Ricinus communis agglutinin. These are readily removed from the 2S albumins by gel filtration under non-reducing conditions since the 2S albumins are considerably smaller in size.

The 2S albumins appear to be composed of five major bands under reducing conditions. No previous reports have stated the number of subunits associated with the 2S albumins since reduction of the mature proteins generally led to their irresolution on polyacrylamide gels buffered with glycine. As previously mentioned, two major castor bean 2S albumins have been reported (Tully and Beevers, 1976) of which one has been

characterised (Sharief and Li, 1982). This protein is resolved here (Figure 27, lane A) into three major polypeptide subunits named LS2, SS2 and SS3. Staining of such gels has proved difficult. As previously reported, the characterised 2S albumin does not silver stain, and when Coomassie blue stain is used either the staining of the small subunits is inefficient or more likely the small subunits are rapidly leached from the gel (McGurl, 1986). As a consequence, the smaller subunits are often difficult to visualise. In order to investigate these proteins further they were blotted onto nitrocellulose by the Western method and probed with rabbit antiserum containing antibodies raised towards the previously characterised 2S albumin. Figure 28 shows that the characterised albumin is represented by LS2, SS2 and SS3 (Figure 28, lane 2) although in total soluble matrix extracts, SS3 is not readily apparent (Figure 28, lane 1). It has been noted that the characterised 2S albumin is composed of a single large subunit and a single small subunit (Sharief and Li, 1982) but when purified runs as a doublet of proteins on polyacrylamide gels (McGurl, 1986 and Figure 29 lane 1). McGurl (1986) noted that the doublet was due to the presence of two large subunits and one small subunit. This contradicts the result obtained here which suggests that the small subunits are responsible for the size heterogeneity.

McGurl suggested a number of possibilities for the presence of two differently sized subunits. Glycosylation was discounted on a number of counts including previous observations (Tully and Beevers, 1976; Roberts and Lord, 1981b), failure to stain with Periodic Acid-Schiff reagent, lack of glycosylation in other 2S albumins and serine protease

inhibitors, and lack of a potential glycosylation site. It was suggested that it could be due to a novel system of conformational flipping in the large subunits, such that they differed when migrating on polyacrylamide gels. This was believed to be due to the high disulphide bond content. This would seem an unlikely explanation if the small subunits were to be implicated in the size difference, since they only possess two cysteine residues which do not disulphide bond with themselves. The most probable explanation, it was suggested, was that the subunit differences were a result of the expression of more than one gene, with differences in the coding regions leading to a truncation at the carboxyl terminus of one of the large subunits. This seems the most likely explanation if the small subunits are to be implicated as responsible for the heterogeneity. However, it is not wise to suggest that the carboxyl terminus is the only place where size heterogeneity would be introduced. Size differences may be due to more than one gene whose products are microheterogeneous in sequence, as is the case of the zein genes in maize (Pederson et al, 1982) and four isoforms of the soybean Bowman-Birk protease inhibitor are known to be almost identical except for the length of their amino termini (Odani and Ikenaka, 1978).

Size heterogeneity of the small subunits could be produced by a post-translational (or even artefactual) proteolytic cleavage step rather than by differences in gene products. The Brazil nut methionine rich albumin has been cloned and well characterised, and it is known that 80% of the large subunits are 3 amino acid residues longer at the amino terminus than the remainder (Altenbach et al, 1987). In comparing the castor bean albumin with the Brazil nut protein, Altenbach et al noted

that the junction between the large and small subunits of the castor bean protein corresponded to the minor cleavage site of the Brazil nut 12 kDa precursor which generated size heterogeneity in the subunits. It was speculated that the castor bean 2S albumin may be synthesised as part of a larger precursor like the Brazil nut and that the final processing step involved in the maturation of the castor bean protein may be similar to that found with the Brazil nut protein. It is worth noting that the tripeptide cleaved in the Brazil nut generating two different large subunit sizes is PRO-ARG-ARG, which is found at the carboxyl terminus of the castor bean albumin small subunit, SS2 (Figure 15, residues 167-169). A similar processing step would then generate two differently sized small subunits consistent with the results found here.

Of the five subunits associated with the 2S albumin fraction, three have been associated with the previously characterised 2S albumin, that is those labelled LS2, SS2 and SS3 (Figure 27), where one molecule of SS2 or SS3 is disulphide linked with one molecule of LS2 to generate the mature protein (note that the subunits are already physically linked by the precursor molecule as well as by disulphide bonds so that processing does not generate free subunits as such).

Figure 29 shows that the two remaining polypeptides named LS1 and SS1 are not recognised by antibodies raised towards the previously characterised 2S albumin. This then suggests that LS1 and SS1 are not differently processed forms of the characterised 2S albumin, but are immunologically distinct. It should be noted also that the polypeptides appear to be equimolar, with LS1 the same as LS2, and SS1 the same as the sum of SS2 and SS3, though this was not determined empirically. It

could be then that LS1 and SS1 represent the subunits of the putative storage protein present on the preproalbumin, since their location is correct (the soluble matrix of the protein bodies), the sizes are as expected (small and large subunits within the 2S albumins), they were found to be unreactive towards silver stain, are immunologically distinct from the previously characterised 2S albumin and are in equimolar amounts (as would be expected if all were generated from a common precursor).

Author's note:
These subunit proteins were individually excised from a polyacrylamide gel using a Schleicher and Schuell Biotrap apparatus and sent for amino terminal analysis by sequential Edman degradation in March 1989. As yet the results of this are unknown (September 1989, SDI).

Finally, attention should be drawn to the activity, during probing of Western blots, of the unreduced and reduced characterised 2S albumin towards rabbit polyclonal antibodies that were raised against it (Figure 29, lanes 1 and 2 respectively). It was necessary to use only 1 ug of unreduced 2S albumin to achieve a similar reaction with 10 ug reduced 2S albumin. This reinforces the findings of McGurl (1986) who also demonstrated this and suggested that the antigenicity of the 2S albumin was heavily dependent on three dimensional structure. In line with the protocols adopted by McGurl (1986) the amounts of the previously characterised 2S albumin were not determined in heterogeneous mixtures due to the lack of a specific assay for the protein.

Figure 30

In vitro processing of preproalbumin

Preproalbumin transcripts generated from clone 14g4 subcloned into pGEM 3blue as described (M3.1.1 and figure 10) were in vitro translated in a wheatgerm system (M2.2.b). 3 ul translation products were incubated overnight at 25°C with either 7 ul citrate/phosphate buffer or 7 ul total soluble protein body proteins (M1.4). The products were resolved along with prestained molecular weight markers on a 15% (w/v) denaturing polyacrylamide gel under reducing conditions (M1.1.a). The gel was stained in Coomassie brilliant blue R, followed by fluorography, drying and exposure to film overnight at -70°C.

Lane 1 3 ul translation products incubated with 7 ul total soluble protein body proteins

Lane 2 3 ul translation products incubated with 7 ul citrate/phosphate buffer

Lane 3 3 ul translation products

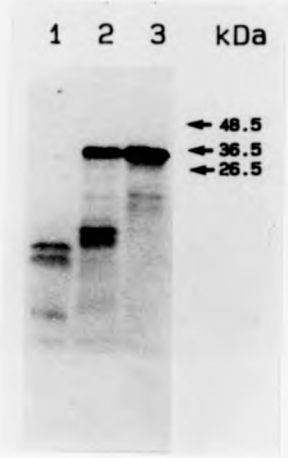

1
2
3
kDa
48.5
36.5
26.5

Figure 31

The time course of processing preproalbumin in vitro.

11.5 ul translation products were incubated at 25°C with 9.5 ul total soluble protein body proteins as described (figure 30). 3 ul aliquots were removed at hourly intervals and placed in reducing sample buffer. After six hours samples were resolved with prestained molecular weight markers on a 16.5% (w/v) denaturing polyacrylamide gel buffered with tricine (M1.1.b). The gel was stained in Coomassie brilliant blue R, fluorographed, dried and exposed to film overnight at -70°C (M2.2.e).

Lane 1	With no incubation
Lane 2	With 1 hour incubation
Lane 3	With 2 hour incubation
Lane 4	With 3 hour incubation
Lane 5	With 4 hour incubation
Lane 6	With 5 hour incubation
Lane 7	With 6 hour incubation


kDa
1 2 3 4 5 6 7
58
48
36
26

D3.4 In vitro processing of the preproalbumin

If the preproalbumin does generate two distinct heterodimeric storage proteins then it might be possible to simulate the post-translational endoproteolytic cleavage by incubating preproalbumin polypeptides generated by an in vitro transcription/translation method with total soluble protein body proteins taken from developing castor bean seeds. This technique using preproricin, immunoprecipitated from in vivo labelled developing endosperm tissue, showed correct processing to the constituent subunits (Harley and Lord, 1985).

Figure 30 shows the result of an overnight incubation of the preproalbumin labelled with S^{35}-methionine with and without the total soluble protein body proteins. The precursor disappears (Figure 30, lane 1) but the bands generated do not represent the known and expected sizes of the characterised and putative 2S albumins.

In order to ensure labelling of each subunit S^{35}-cysteine was employed in the in vitro translation. The incubation was also placed under timed conditions and the products resolved on a tricine buffered polyacrylamide gel. Figure 31 shows the results of such an experiment. It is clear that the products are rapidly cleaved from the precursor over a period of hours to generate smaller fragments. Once again, these fragments are not the sizes expected of mature 2S albumin subunits. It would also seem as though some of the processed peptides were

accumulating, but these were not readily apparent as major proteins when total soluble castor bean proteins were stained with Coomassie blue.

Proricin and proRCA have been shown to generate their constituent subunits by both *in vivo* and *in vitro* proteolytic processing (Butterworth and Lord, 1983; Harley and Lord, 1985). In both cases it was noticed that a precursor of 34kDa (identified as the preproalbumin) was N-terminally cleaved to remove a signal sequence upon entry into the endoplasmic reticulum, and thereon became post-translationally processed. The products of processing were not of the correct size for the mature protein and indeed closely resembled the result shown in Figure 30 (lane 1). It would seem then that the post-translational processing of the preproalbumin is complicated and has not yet been resolved.

So far no enzymes have been purified that are involved in the post-translational processing of storage proteins. Altenbach *et al* (1986) showed that the sulphur rich seed protein from Brazil nut was cleaved rapidly from the full length precursor of 18 kDa to 12 kDa through a 15 kDa intermediate. The 12 kDa polypeptide was then subsequently accumulated in the developing seeds over a period of 8-9 months after flowering. It was then processed very slowly resulting in the 9 kDa and 3 kDa mature subunits. De Castro *et al* (1987) showed that this final process in an *in vivo* pulse-chase experiment required more than 12 hours. In defining the processing sites of the *Arabidopsis thaliana* 2S albumins, Krebbers *et al* (1988) suggested that as many as four different proteases were required to process the proalbumin after signal cleavage.

The incomplete (or artefactual) processing of the preproalbumin shown in Figures 30 and 31 could be a result of only some of the required processing enzymes being present, or a failure to leave the incubation long enough for the action of a very limiting or slow protease to manifest itself. It seems unlikely that processing of the preproalbumin into its predicted mature subunits will be a simple matter but will require a more thorough and exhaustive approach.

D3.5 Conclusions

The protein bodies of the castor bean endosperm contain 2S albumins that are composed chiefly of five polypeptides, three of which are attributed to the previously characterised 2S albumin. The other two could represent the subunits of the putative 2S albumin. The heterogeneity of the previously characterised 2S albumin, manifested in the presence of a doublet of proteins, is due to the presence of two differently sized small subunits which may differ by a post-translational processing step at their carboxyl termini. *In vitro* processing of the preproalbumin by protein body extracts failed to generate the mature subunits. The preproalbumin shares no antigenic determinants with the prepronapin of *Brassica napus* when both are generated *in vitro* in a reducing environment.

D4 ANALYSIS OF THE PREPROALBUMIN GENES

D4.1 Summary of Results

Total RNA extracted from castor bean seeds at different stages of development was electrophoresed on a formamide/agarose gel and transferred to nitrocellulose by Northern blotting. After probing with preproalbumin cDNA clone 14g4, hybridisation signals were apparent from RNAs extracted from seeds at stages of development D, E and F, that is from seeds where testa formation was apparent to seeds that were desiccating. Signals were not apparent from seeds before the onset of testa formation, from mature dry seeds, nor from germinating seeds.

Castor bean genomic DNA that was restriction endonuclease digested and electrophoresed on an agarose gel was transferred to nitrocellulose by Southern blotting. On probing with clone 14g4 hybridisation signals were apparent from at least four fragments with differences noted in the intensities of each signal.

A preexisting castor bean genomic library was screened with clone 14g4 and ten positively hybridising plaques were purified. DNAs isolated from these phage were digested with restriction endonuclease Eco R1, electrophoresed on an agarose gel and transferred to nitrocellulose by Southern blotting. On probing with clone 14g4 a number of different clones were identified, some of which contained more than one hybridising Eco R1 fragment. The fragment from clone 6a and the higher and lower length fragments from clone 5 which all positively hybridised to clone 14g4, were subcloned into vector puc19x to generate subclones

pgcep6a, pgcep5h and pgcep5l respectively. The restriction maps of the subclones were identified and found to be different.

Sequence data were generated from all three subclones. 1,591 bases of sequence were generated from clone gcep6a, spanning the entire length of cDNA clone 14g4 plus 470 bases upstream of the translation start point. 489 bases of sequence were generated from clone gcep5l including 370 bases upstream of the translation start point. Clone gcep5h was incomplete but positive identification was made by sequencing. Clone gcep6a was directly comparable to the cDNA sequences obtained and contained no introns.

The upstream regulatory regions of clone gcep6a were compared with those of gcep5l, and the those of the genes encoding ricin, the Brassica napus napin and the Arabidopsis thaliana 2S albumin. Homologies were discovered between the albumins in a highly conserved region spanning approximately 70 bases immediately upstream of their TATA boxes. Homology with the ricin gene was limited to a 12 base region located within this highly conserved area.

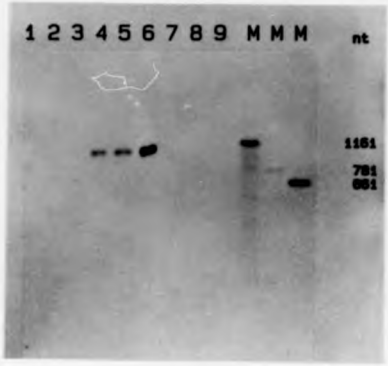
Figure 32

Northern blot of castor bean seed RNA extracted from different developmental stages probed with cDNA clone 14g4

Total RNA from different stages of seed development was electrophoresed on a 50% formamide agarose gel (M2.3.b) and blotted onto nitrocellulose as described (M2.4). The filter was probed with labelled cDNA clone 14g4 and subjected to high stringency washing. The filter was exposed to film for eight hours at -70^{0}C.

Lane 1 10 ug developing seed RNA: stage A
Lane 2 10 ug developing seed RNA: stage B
Lane 3 10 ug developing seed RNA: stage C
Lane 4 10 ug developing seed RNA: stage D
Lane 5 10 ug developing seed RNA: stage E
Lane 6 10 ug developing seed RNA: stage F
Lane 7 10 ug dry seed RNA
Lane 8 10 ug germinating seed RNA: 3 days growth
Lane 9 10 ug germinating seed RNA: 7 days growth

Stages of seed development are as defined by Roberts and Lord (1981a). Size marker transcripts run were of clone 14g4 (1161 nucleotides), napin clone pN2 (781 nucleotides) and alpha factor (661 nucleotides). These were generated from T7 polymerase transcription of the clones in vector pGEM 3blue. RNA samples were a generous gift from J. Tregear.


1 2 3 4 5 6 7 8 9 M M M
nt
1161
781
661

D4.2 Developmental Regulation of Preproalbumin Transcription

Total RNA extracted from the seeds of castor bean plants harvested at different stages of development were obtained form J. Tregear (Warwick, UK). Harvesting had been performed as described (Roberts and Lord, 1981a). Stages A, B and C represented small, medium and large sized seeds prior to testa formation, stage D was when testa formation was observed to have begun, at stage E the testa was formed and at stage F the seeds were desiccating. Dry mature seeds were used as well as germinating seeds harvested 3 and 7 days post imbibition.

Upon electrophoresis of the RNAs in a 50% formamide/agarose gel and transfer to nitrocellulose by Northern blotting, the blot was probed with clone 14g4 and washed at high stringency. This gave the result shown in Figure 32. A single RNA band appears to hybridise to the cDNA probe, with a length of approximately 1.0-1.1 kb. This is similar to the result obtained in Figure 9 where poly A^+ RNA extracted from post-testa seeds was probed with clone 14g4. The hybridising RNA only appears in stages D, E and F, that is with the onset of testa formation until the seed is desiccating. Once the seed has desiccated to the mature dry seed no hybridising RNA is seen, nor is it seen in germinating seeds. This suggests that the hybridising RNA is a transcript that is regulated during the development of the maturing castor bean seed.

It can be seen that there is a greater hybridising signal in RNA from desiccating seeds (Stage F: Lane 6, Figure 32) than in RNA from stages D and E (Lanes 4 and 5, Figure 32). This could be due to a greater number

of transcripts generated by higher gene expression during desiccation, or perhaps may reflect a difference in the proportion of poly A^+ RNA to the total RNA loaded onto the gel. Repetition of the experiment using equivalent amounts of poly A^+ RNA would clarify this position. Size markers generated by *in vitro* transcription of cDNAs subcloned into pGEM 3blue were longer than the cDNAs themselves due to the presence of part of the multiple cloning site located at the 3' end of the cDNA (Figure 10).

The expression of plant storage protein genes is regulated with respect to both time and place: most seed storage proteins are expressed only or largely in the seed and only at certain stages of seed development (Chandler *et al*, 1984; Crouch and Sussex, 1981; Walling *et al*, 1986; Altenbach *et al*, 1987; Casey *et al*, 1986; Laroche-Raynal and Delseny 1986; reviewed Higgins, 1984).

It has been demonstrated recently that ricin is developmentally regulated in a similar way (Tregear, 1989) and is also tissue specific, not being found in stems, leaves or roots. In castor bean seeds ricin is generated at a similar time to the 2S albumins and both are deposited within the protein bodies of the castor bean endosperm during post-testa seed development (Roberts and Lord, 1981b; Butterworth and Lord, 1983; Lord 1985ab). The data generated here suggest that the preproalbumin and preproricin transcripts are generated at a similar stage of seed development. It could be then, that the genes generating these transcripts are regulated in a similar manner, perhaps by a trans-acting factor that binds to the upstream regulatory regions. In this case these regions should possess some homology to each other. Ricin genes have

already been cloned (Halling et al, 1985; Tregear, 1989) and it seems timely that a genomic clone for the preproalbumin be obtained since a cDNA probe now exists, with a castor bean genomic library in existence that successfully generated a ricin genomic clone in this laboratory (Tregear, 1989). The cloned preproalbumin gene could also be compared with other 2S albumin genes already cloned (Krebbers et al, 1988; Scofield and Crouch, 1987; Josefsson et al, 1987; Allen et al, 1987; Higgins et al, 1986).

Figure 33

Southern blot of restriction enzyme digested castor bean genomic DNA probed with cDNA clone 14g4

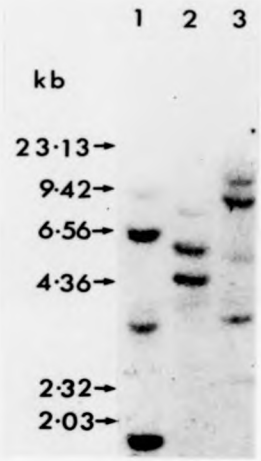
10 ug aliquots of castor bean genomic DNA were restriction endonuclease digested with Bam HI. Hind III and Eco RI (M3.1.f). The DNA was electrophoresed on a 0.7% (w/v) agarose gel buffered with TBE (M3.1.e) and blotted onto nitrocellulose which was then probed with cDNA clone 14g4 (M3.4). Following high stringency washing the filter was exposed to film for 48 hours at -70°C.

Lane 1 DNA digested by Bam HI

Lane 2 DNA digested by Eco RI

Lane 3 DNA digested by Hind III

Size markers were lambda DNA digested with Hind III. Castor bean genomic DNA was a kind gift of J. Tregear.


1
2
3
kb
23·13
9·42
6·56
4·36
2·32
2·03

D4.3 Estimation of the preproalbumin gene copy number

Genomic DNA previously extracted from castor beans (Tregear, Warwick, UK) was subject to restriction endonuclease digestion with enzymes Hind III, Eco R1 and Bam H1. The fragments generated were electrophoresed in an agarose gel and transferred to nitrocellulose by Southern blotting. After probing with clone 14g4 the filter was washed at high stringency. The result is shown in Figure 33. The enzymes do not possess restriction sites within the cDNA clones obtained, so without ruling out the possibility of introns or non-expressed gene members containing such sites it would clearly seem that the preproalbumin is represented by at least four genes. Some of these genes hybridise more strongly to the probe than others suggesting that the gene family is heterogeneous. This was suspected from restriction analysis of the cDNA clones obtained (Table 1).

Most other plant storage protein genes so far isolated are members of gene families, well reviewed in the somewhat dated article by Higgins (1984). More closely, the 2S albumin family of proteins appears also to be encoded by gene families. In sunflower, the albumin is represented by at least two divergent genes (Allen et al, 1987). In Arabidopsis thaliana, four different albumin gene members have been cloned (Krebbers et al, 1988). In Brassica napus the napin has been seen to be represented by a gene family in the order of a minimum of ten members and possibly as many as sixteen (Josefsson et al, 1987; Scofield and Crouch, 1987). In pea, at least four different genes for the seed

albumin are expressed in the genome (Higgins *et al*. 1986). All this reflects the belief that plant storage proteins in general are members of a superfamily and originated from an ancestral gene (Kreis *et al*. 1985a), but have retained their homology because there may be a need for a fairly tightly constrained protein structure which must fulfil certain criteria (Spencer, 1984). The preproalbumin from castor bean appears to be no exception.

Figure 34

Screening the castor bean genomic library

The castor bean genomic library, created by J. Tregear, was screened with the cDNA clone 14g4. An example of the hybridisation signals obtained from the first filters is marked 'A'. A plaque that was picked from this plate is indicated with an arrow. This plaque was purified by two more rounds of screening until when plated and screened, all plaques present hybridised to the probe to give the result marked 'B'.

Facing -240-

A

B

Figure 35

Southern blot of Eco RI digested DNA prepared from genomic clones that positively hybridised to the cDNA clone

DNA was prepared from the plaque pure genomic clones (M3.1.k) and 2 ug aliquots were digested with restriction endonuclease Eco RI (M3.1.f). The fragments were electrophoresed on a 1% (w/v) agarose gel buffered with TBE (M3.1.e). The DNA was transferred after visualisation and photography to nitrocellulose and was probed with cDNA clone 14g4. After high stringency washing the filter was exposed to film overnight at -70°C.

Part A shows the gel visualised before blotting and part B shows the autoradiograph of the probed filter.

Lane 1	Clone 1
Lane 2	Clone 2
Lane 3	Clone 4
Lane 4	Clone 5
Lane 5	Clone 6b
Lane 6	Clone 11
Lane 7	Clone 12a
Lane 8	Clone 12b
Lane 9	Clone 13
Lane 10	Clone 6a

Size markers were the restriction fragments of lambda DNA digested with Hind III.

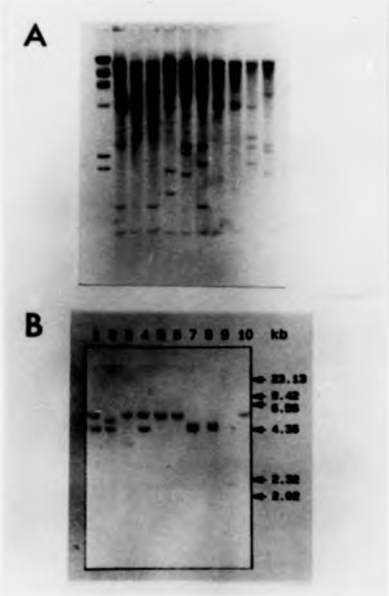

A
B
1 2 3 4 5 6 7 8 9 10 kb
23.13
9.42
6.55
4.36
2.32
2.02

Figure 36

Restriction enzyme maps of the subcloned genomic clones

Eco RI fragments of genomic clones 6a and 5 that positively hybridised to the cDNA clone were subcloned into puc19x. Their restriction endonuclease sites were then determined and are shown here. Enzymes applied to the clones include: Snh I: Pvu II: Hind III: Sac I: Ava II: Kon I: Cla I: Pvu I: Eco RV: Pst I: Sal I: Hinc II: Xba I: Bam HI: Bgl II; Nco I: Sma I: Xho I.

The green hatched area delineates where the cDNA clone was found to be encoded. Clone 5 had two fragments subcloned: Clone 5L (Light) was 4.3kb long whilst clone 5H (Heavy) was similar in length to clone 6a, 5.4 kb.

5H
EcoR1
Sph 1
Sac 1
Pvu 2
EcoR1
Sph 1
Sac 1
Hind 3
EcoR1
6000p
5L
EcoR1
Sph 1
Sac 1
Pvu 2
Hind 3
Sph 1
Sac 1
Sph 1
EcoR1
6A
EcoR1
Hind 3
Sph 1
Sac 1
Pvu 2
Hind 3
Sph 1
Sac 1
EcoR1

D4.4 Cloning of the preproalbumin genes

A genomic library had been created using castor bean genomic DNA that was partially digested with restriction endonuclease Sau 3A, size fractionated for fragments of 12 kb by sucrose density gradient centrifugation, ligated into phage lambda vector Charon 35 (Loenen and Blattner, 1983) restriction endonuclease digested with Bam Hi and *in vitro* packaged. Infection of *E.coli* K803 was used to generate plaques from which was isolated a ricin clone (Tregear, 1989).

This library was screened with clone 14g4. $5x10^5$ plaques were probed from which approximately 40 gave strong positive hybridisation signals, indicating a small genome for the castor bean, as previously noted by Tregear (1989). 10 of these plaques were carried forward by sequential rounds of purification and screening to plaque purity (Figure 34). Filters were washed at high stringency so it was possible that less closely related genes were not represented.

DNAs were extracted from minipreparations of lambda clones. These were subject to Eco RI restriction endonuclease digestion followed by electrophoresis of the fragments on an agarose gel (Figure 35, part A) and transfer to nitrocellulose by Southern blotting. Probing of the blot with clone 14g4 produced the result shown in Figure 35, part B. Some of the clones were obviously the same with similar restriction patterns, for example clones 12a and 12b (Figure 35, lanes 7 and 8), clones 6a and 6b (Figure 35, lanes 10 and 5). Some of the clones appeared related in terms of the size of their hybridising insert(s), but with dissimilar

restriction patterns, for example clones 1 and 5 (Figure 35, lanes 1 and 4), clones 4 and 11 (Figure 35, lanes 3 and 5) with clones 6a and 6b (Figure 35, lanes 10 and 5). This difference was probably brought about by the low specificity of the restriction endonuclease Sau 3A cutting incompletely as required. This would generate clones with overlapping fragments, with the gene located on a defined Eco R1 fragment, but with different length 5' and 3' ends. Where only part of the clone was represented in the phage DNA, because of Sau 3A cleavage within the gene itself (there are 6 sites located in cDNA clone 14g4), hybridising Eco R1 restriction fragments would be smaller than expected, as may be the case with the upper bands of clones 2, 12a and 12b (Figure 35, lanes 2, 7, 8). The hybridising insert of clone 13 (Figure 35, lane 9) was much smaller than the others (approximately 2.25 kb)

Some clones contained two hybridising inserts. This was not thought to be due to incomplete digestion because of the equal quantities of both bands and because digestion with excess enzyme made no difference. The two bands could be produced from a gene containing an Eco R1 site located within the transcribed region (but this was not seen in any of the cDNA clones) or located within an intron. It could also be due to the presence of two preproalbumin genes closely linked to each other. The four Arabidopsis 2S albumin genes are all tightly linked in a tandem array (Krebbers et al. 1988).

The hybridising fragment from clone 6a and the higher and lower hybridising fragments from clone 5 were subcloned into puc19x, generating the pgcep6a, pgcep5h and pgcep5l plasmids. The restriction maps of these clones were determined and are indicated in Figure 36. It

was apparent that the maps of all three clones were different, hence three different genes appeared to have been cloned. Comparison of the restriction maps of the genomic and cDNA clones determined the position of the genes within the subcloned genomic fragments. Clone gcep5h was incomplete by truncation at the 5' end, and appeared to be lacking the 3' Pvu II and Sac I sites. Nevertheless its identity was confirmed by limited 5' sequencing (see D4.5). This suggested that different genes present on clones gcep5l and gcep5h were linked on the same genomic fragment in clone 5. Clones gcep6a and gcep5l were different in size. By possession of the 3' Sac I site however, they both resembled cDNA clone 14g4 more than clones 8g8 or 10a12 (assuming that the difference between the cDNA clones is not a function of the use of a different polyadenylation signal as discussed in subsection D2.3).

In the light of the restriction maps, when the clones are compared with the Southern blot of Eco R1 restriction endonuclease digested castor bean genomic DNA (Figure 33, Lane 2), it would seem that clone gcep5l represented the fragment of length 4.3kb, whilst clone gcep6a represented the clone of length 5.4 kb. It is suspected that clone gcep5h either represents a fragment of similar size to clone gcep6a, which is why only three bands are seen to hybridise in Eco R1 digested castor bean DNA instead of four, or that it represents the larger fragment of approximately 8.1 kb, but is incomplete due to Sau 3A digestion. This would agree well with the observation that the 8.1 kb fragment does not hybridise as strongly to the clone 14g4 as do the fragments of 4.3 and 5.4 kb, which may be a reflection of the difference in the restriction maps seen between gcep5h and clone 14g4. Hence gcep5h

could encode for a variant member of the castor bean preproalbumin gene family.

Clones gcep5i and gcep6a appeared to possess no introns when their maps were compared with cDNA clone 14g4. Whilst an intron has been found in the sunflower albumin, none have been seen in the Brassica napin genes or the pea seed albumin gene so far characterised, or in any of the Arabidopsis 2S albumin genes. (Allen et al, 1987; Scofield and Crouch, 1987; Josefsson et al. 1987; Higgins et al. 1986; Krebbers et al, 1988). Scofield and Crouch (1987) stated that lack of introns was characteristic of genes for many other 2S seed proteins and related cereal prolamins.

Figure 37

Sequencing strategy for genomic subclones.

The areas of clones 6a, 5h and 51 that were subject to sequencing are indicated. The extent and the direction of sequencing are indicated by arrows underneath each subclones' restriction map.

Facing -247-

EcoR1
Sph 1/Hind 3
Sac 1
Sph 1
Pvu 2
Pvu 2
Sac 1
Hind 3
6A
EcoR1
Sac 1
Kpn 2
Pvu 2
Sph 1
Sac 1
Sph 1/Hind 3
Pvu 2
5L
EcoR1
Sph 1
Pvu 2
Sph 1
5H
600bp

Figure 38

DNA sequence of genomic clone gcep6a compared with cDNA clone 14g4

The sequence of genomic clone gcep6a is listed and numbered from the start. A comparison is made with cDNA clone 14g4 with only the differences indicated. Asterisks indicate two bases in clone gcep6a not present in clone 14g4. The limits of sequence for cDNA clone 14g4 are shown by square brackets.

Facing -248-

```
                                ATAAGCAATGATTTTATCTAGCCTGGTATATGTCCGCAGTATA    43   gcap6a
TACTAAGTTGATAGATGACTGATACATATTTTATTAATTTTAAATAATAAAATATATATT   103   gcap6a
AACTACCTAATATTAAAATATAAAATATTTATATATATTTCATACTTTTATTAATAAAAA   163   gcap6a
AAATATAATTATTATCAAATGTACATAAAAATAAAATAAATTTAAAAATATATACTTCAA   223   gcap6a
CAACATCAAGCGCAAAATCCATAAAAAGAATCATATGAAGCAAACAGAAACTTGCATGCT   283   gcap6a
GCTGCCATACGTGTAAGCTTAAAGCAGACCTCAAAACCATGGGTAGCCACCCCATAAAAA   343   gcap6a
TGCCAGAAATGACTGGCCATTCCATACATCTTACACGTTCTCGACATGCACTCCCCTAAA   403   gcap6a
TTTCCTATAAGAGCCCATTACCATCATCTACTATATTCCATAAGAGCATCTTTCTGATAT   463   gcap6a
TCAGTAACAATGGCAAAGCTCATACCCACAATTGCTCTTCTTAGTGTTCTCTTGTTCATC   523   gcap6a
[  A                                                      T          14g4
ATTGCCAATGCGTCTTTCGCTTACAGGACCACCATCACCACCATTGAGATTGACGAGTCA   583   gcap6a
                                                                     14g4
AAGGGTGAAAGGGAAGGATCGAGCTCGCAGCAATGCCGCCAGGAAGTTCAGAGGAAGGAC   643   gcap6a
                                                                     14g4
TTCAGCTCCTGCGAGCGATACCTGAGGCAATCAAGTTCAAGAAGATCACCAGGAGAAGAA   703   gcap6a
                                                          A          14g4
GTGTTAAGGATGCCTGGAGATGAAAACCAGCAGCAGGAGAGCCAGCAACTCCAGCAATGC   763   gcap6a
                                                                     14g4
TGCAATCAGGTAAAGCAAGTAAGAGATGAATGCCAATGTGAAGCAATCAAATATATCGCA   823   gcap6a
                                                                     14g4
GAGGATCAGATTCAGCAGGGACAGCTACATGGAGAAGAGTCTGAAAGAGTGGCGCAGAGA   883   gcap6a
                                                                     14g4
GCAGCTGAAATTCTATCTTCTTGCGGTGTGCGTTGCATGCGCCAAACTCGAACAAACCCA   943   gcap6a
                                                                     14g4
AGCCAGCAGGGGTGTGGTGGGCAGATTCAAGAGCAACAAAATCTCAGGCAATGCCAGGAA  1003   gcap6a
                                                                     14g4
TATATCAAACAACAAGTTTCCGGACAGGGACCAAGAAGAACTGACAATCAAGAACGCTCT  1063   gcap6a
                                                                     14g4
CTTCGTGGGTGCTGTGACCATCTCAAGCAGATGCAGTCACAGTGCAGATGCGAGGGTCTC  1123   gcap6a
                                                                     14g4
AGGCAGGCTATTGAGCAGCAACAGAGCCAGGGGCAACTTCAAGGTCAGCATGTTTTTGAG  1183   gcap6a
                                                                     14g4
GCTTTCAGGACAGCTGCGAATTTGCCATCAATGTGCGGCGTCTCACCAACCGAATGCCGG  1243   gcap6a
                                                                     14g4
TTCTAAAGTCGCAGGATTATATATATATATGAAGTAGTTAATAAAGATCATGATCAGTGA  1303   gcap6a
                                **                                   14g4
TCACTAGCTCTTCGCAGCTGTTGCTGTGAGATTAGAAAATGATGATTACGTAGTAGCCAT  1363   gcap6a
     C                                                               14g4
GTATCATCACTGTTGCAAAATCTATAATAAGTTTATGTGGCTCCTCTGTTAATTAATATC  1423   gcap6a
                                                                     14g4
TCATGAGCCTACCCAACTCCATTAACGTAGGACTGGAGCTCAAACTTTAGGACTCAGGAC  1483   gcap6a
                                                                     14g4
TCAGAGACTGGTACTTTAACTGTATTAGAACTTTGTTAAGGAAAATAAATAACTTAAAAG  1543   gcap6a
                                                                     14g4
AGCAAATGCGAAAGCCCCAGTGTATTCAAGATTACTCTGGCATTTCAG              1591   gcap6a
          polyA]                                                     14g4
```

D4.5 Sequences of preproalbumin genomic clones

The areas which showed restriction maps similar to the cDNA clone 14g4 were subject to DNA sequencing by subcloning fragments into vectors M13mp18 and M13mp19 and using the dideoxynucleotide chain termination method of Sanger et al (1977), modified by the United States Biochemical Corporation (1988) for use with Sequenase which gave better results than using the Klenow fragment of E.coli DNA polymerase 1. The sequencing strategies of all three clones are indicated in Figure 37. Clone gcep6a generated 1,591 bases of sequence spanning the entire sequence generated by cDNA clone 14g4 with an extra 470 bp upstream of the translation start point. Clone gcep5l generated 489 bases of sequence including 370 bp upstream of the translation start point. Clone gcep5h was positively identified as an albumin clone only, but was identical to clone 14g4 in the 260 bp of sequence generated at the 5' end.

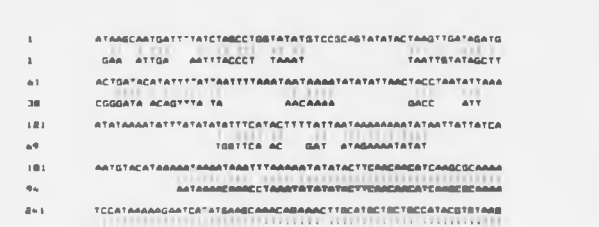
The sequence from genomic clone gcep6a was compared with cDNA clone 14g4. The result is shown in Figure 38. Whilst the restriction maps of both clones suggested they were similar there are a number of base changes and insertions. Whilst the stretch of sequence that encodes the signal sequence more closely resembles clone 14g4 than clone 8g8, the substitution of base C for base A at residue 691 (Figure 38), and the insertion of an extra AT (bases 1261-1262) in the A/T repeat region at the 3' end of the clone suggest similarities to clone 8g8, a group 1 cDNA clone that does not possess the 3' Sac I site. The presence of the 3' Sac I site in both clones gcep6a and gcep5l suggests that the

difference noted between group 1 and 2 cDNA clones is an artefact generated by the use of different polyadenylation signals noted at the end of the clones (see subsection D2.3). Isolation of all gene members would clarify the situation. Base changes were also noted at positions 512 and 1312 not present in either clone 8g8 or 14g4 suggesting that clone gcep6a does not represent either cDNA clone but a very homologous member of the family.

Figure 39

Comparison of the sequence of genomic clones gcep6a and gcep51

The sequence obtained for genomic clone gcep51 (1-489 bases) was compared with part of the sequence obtained for clone gcep6a (1-590). Gaps were introduced to increase homology. The top row of each comparison is the sequence of gcep6a, with the bottom row that of gcep51. The translation start point is numbered 473 in clone gcep6a.



Matches = 662 Mismatches = 87 Unmatched = 101
Length = 850 Matches/length = 78.3 percent

Figure 40

Comparison of the regulatory regions of the ricin gene and the albumin gene

The regions upstream of the translational initiation codon were compared between the gene encoding preproricin (gric) and the gene encoding the preproalbumin (gcep6a). The regions are compared from the initiation codon backwards (Strand L) with gaps introduced to achieve a better homology. The upper sequence is the preproalbumin sequence (limits 162-473) whilst the lower is the preproricin sequence (limits 1-311). The sequence of the ricin gene was taken from Tregear (1989) and the numbering conserved. Highlighted areas include the 'TATA' box of the ricin gene and a region of homology shared between the two sequences outside of the linear comparison.

```
gcep6a                                  STRAND L
gRic                                    STRAND L

473          TTG TTACTGAATATCAGAAA   GATGCTCTTATGGAATATAGTAG ATGATGGTA
             ||| || | | |  || | ||    ||   ||     |||| | ||| ||  |  ||
311      TCTTTGATTGCAGCAATTCTGGAATTTTATTGGCTGTCCAAATACATTAGCATTCTCATA

421      ATG GGCTCTTATAGGAAATTTAGGGGAGTGCATGTCGAGAACGTGTAAGATGTATGGAA
         | |  |  ||||| |   ||  | |  ||       | |   | | | ||  | |  ||
251      AAGATGAACTTATAAAATTAATACAGAAAAACACTAATACATTTTTTTATAT AAAGAAA

362      TGGCCAGTCATTTCTGGCATTTTATGGGGTGGCTACCCATGGTTTTGAGGTCTGCTTTA
            ||   | ||     ||   ||    |   |  ||      || | |||  ||
192      ACCACAAATACTTGATCAATAAAAATAACTTATAAATTAATACGGAAGATGCATGCCATA

302      AGCTTACACGTATGGCAGCAGCAT  GCA AGTTTCTGTTTGCTTCATATGATTCTTTTT
            |    |  | |  ||||  |||  ||||       || || || ||
132      TATAAATTGAGACAGAATATGCATAAGCAGTGTTTAAAGGGAATTGAT TGGTTTAACGA

245      ATGGATTTTGCGCTTGATGTTGTTGAAGTATATATTTTTA AATTTATTTTATTTTTATG
         |  ||| |   | ||   |   |||  | |||  || ||| |  |||   | ||
73       AATAATTATAAATTAGAAACGGAATTAGTTAAAATTCATATAATGTCGATTAGAGTAATT

186      TACATTTGATAATAATTATATTTTT
            ||| |||  ||
13       TACTTTTTTTAGG

Matches = 137    Mismatches = 161    Unmatched = 27
Length = 325     Matches/length =  42.2 percent
```

D4.6 Comparison of the regulating regions of preproalbumin gcep6a with those of gcep5l and the ricin gene.

The sequence generated for clone gcep6a was compared with the limited sequence obtained for gcep5l. This is shown in Figure 39. Throughout the open reading frame identified from cDNA clone 14g4 the clones were exactly identical (up until the 3' end of the limited sequence generated for clone gcep5l). The sequences were also identical for a further 262 bases upstream of the translation start point except for one base deletion in clone gcep5l (base 274, clone gcep6a, Figure 39). From this point the two genes diverged rapidly with the only homologies observed occurring against the very high A/T rich region present in clone gcep6a. No further homologies were apparent. Clone gcep5l did not possess such a high A/T rich region.

When the sequence of the upstream regulatory region containing the A/T rich sequence was compared with the NBRF/Genbank databank of sequences using the Microgenie program, the homologies obtained were all to the 3' untranslated regions of genes, suggesting possibly that clone gcep6a is linked in tandem to another gene. This would not leave much distance between the 3' end of one gene and the 5' end of clone gcep6a. It could be that all the genetic information required for tissue and temporal regulation of the gene occurs within the short high area of homology seen between clones gcep6a and gcep5l (Figure 39). Rather than being the 3' end of another gene the A/T rich region could be required for proper transcription initiation. A high A/T rich region approximately 250 bases

upstream of the translational start point was also shown in the Brassica napin genes by Josefsson et al (1987) and Scofield and Crouch (1987). The former group suggested that hairpin loop structures (and possibly cruciform structures) occurred here, and this along with the A/T richness could cause perturbation of the DNA structure.

Figure 40 shows the best alignment generated when the lower strand of clone gcep6a was compared with the lower strand of the ricin genomic clone (Tregear, 1989), from both of their translation start points backwards. This was used as the start of comparison because the lack of S_1 nuclease protection experiments prevented the identification of the transcription start point in the preproalbumin gene. Few areas of significant homology occurred in the direct alignment except around the TATA box of the ricin clone highlighted in Figure 40. An area of homology was located outside of the direct linear comparison and is also highlighted in Figure 40. With the introduction of two gaps, an area of 14 bases is conserved (with one mismatch) which includes the 5' CATGCAT 3' consensus motif (lower strand reads 3' GTACGTA 5') in the ricin sequence. This heptamer motif was characterised by Dickinson et al (1988) as being present in the 5' flanking regions of many legume storage protein genes (lectins, legumins, vicilins and albumins) in single or multiple copies, but absent upstream of non-storage protein genes. Thus it was implicated in the tissue specific and temporal regulation of transcription of these genes. As will be seen later (Figure 43), this motif is located in an area rich in sequence homologies found between 2S albumin storage protein gene regulatory regions. The motif is not entirely conserved in the gcep6a sequence but

is represented by CATACAT in the conserved position or CATGCAC 16 bases downstream. The motif was not conserved in the Arabidopsis 2S albumin gene AT2S1 either (Krebbers et al. 1988).

Considering the similarity between the expression and location of both the 2S albumin and ricin it is surprising that higher homologies were not found between the regulatory regions. Perhaps the 14 bases of conservation found here is all that is necessary for correct temporal and tissue specific expression.

Figure 41

Comparison of the upstream regulatory regions of the napin gene from Brassica napus and the preproalbumin gene.

The regions upstream of the translation initiation codon were compared between the prepronapin gene, labelled 'napA' and the preproalbumin gene labelled 'gcep6a'. Gaps are introduced to maximise homology. Highlighted areas include the 'TATA' box and the repeated alternating purine/pyrimidine blocks of the napin gene, as described by Josefsson et al (1987). Numbering is as described by the authors.

```
gcmp4a
mapA

1          ATAAGCAATGATT TTATCTAGCCTGGTATATGTCCGCGBT ATATACTAAGTTGATAGA
           |  |  |  |||| |  | |  | | |    |  | |  |  |  ||  |||| |
450        AACAGGGACGGGATYCCGAGTTTGGGAGGTTTTGACTCAAATCCCCAATTTATATTGACCB

59         TBACTGATACATATTTTA TT AATTTTAAATAATAAAAT ATATATTAACTACCTAATA
           ||||| |  ||  ||| | ||  ||  |    | ||| |   |  ||| | || ||
709        TBACTAAATCAACTTTAACTTCTATAATTCTGATTAAGCTCCCCAATTTATATTCCCAACB

116        TTAAAATATAAAATATTTATATATATTTCATACTTTTATTAATAAAAAAAAATATAATTAT
             |  |  |  |||||||  |  | ||||  |  || | ||  ||  |  | || |  | |
769        BCACTACCTCCAAAATTTATAGA CTCTCATCCCCTTTTAAACCAACTTAGTAAACGTTT

176        TATCAAATGTACATAAAAATAAAATAAATTTAAAAATATATACTTCAACAACATCAABCB
           | |    | || ||   |  ||  ||| ||| |  | | |  ||  || || |  |||   |
828        TTT  TTTTTA ATTTTATGAAGTTAAGTTTTTACCT TGTTTTTAAAAABAATCGTTCA

236        CAAAAT CCAT AAAAGGAATCATATGAAGCAAACABA AACTTGCA  TGCTG    CT
           || || ||||   | |  || |  |  |    ||| | | | |||   || |     |
884        TAAGATBCCATBCCAGAACATTAGCTACACGTTACACATABCATBCAGCCBCBBAGAATT


287        BCCATACGTGTAAGCTTAAAGCAGACCTCAAA ACC ATGGGTAGCCACCCCATAAAAAT
           |  | | |    ||| |  | ||||| ||| | | |  |  | | || |  |
944        GTTTTTCTTCBCCACTTGTCACTCCCTTCAAACACCTAAGAGCTTCTCTCTCACAGCACA

345        BCCAGA AATGACTGGCCATTCCATACA TCTTACACGTTCTCGACATGC ACTCCCCTA
             || | ||| ||  ||   | ||| || | |||||||  ||| |||| | || |||
1004       CACATACAATCACATGC  GTBCGTBCATTATTACACGTGATCBCCATGCAAATCTCCTT

402        AATTTCCTATAA  GAGCCCATTACCATCA TCTACTATATTCCATAAGAACATCTTTCT
            ||  |||||||  | | |||  | ||| |||    |  |  ||  |||| |
1068       TATAGCCTATAAATTAACTCATCCGCTTCACTCTTTACTCAAACCAAAACTCATCAATAC

459        GATATTCAGTAACAA
             |  | ||| ||
1122       AAACAAGATTAAAAACATACACGA

Matches = 247     Mismatches = 218     Unmatched = 39
Length = 504      Matches/length =  49.0 percent
```

Figure 42

Comparison of the upstream regulatory regions of the gene encoding the Arabidopsis thaliana 2S albumin and the preproalbumin gene

The upstream regulatory regions of the Arabidopsis thaliana 2S albumin gene AT2S1 (Krebbers et al, 1988) and the preproalbumin gene gcep6a were compared up to the translational initiation codon. Numbering is as described by the authors. Gaps are introduced to maximise homology. The putative 'TATA' box of the AT2S1 gene is highlighted.

```
gcmp4a
gar4b2

73          TTTATTAATTTTAAATAATAAAATATATATTAACTACCTAATATTAAAATAT AAAATAT
                 | | ||   |   |                 |   ||  |||| ||  || ||   || |
1           GACAAT ATCCAATGGACCGGTGACCTGCGTGTATAAGTAATTTTTCAAGATGCTAAAAC

132         TTATATATATTTCATACTTTTATTAATAAAAAAAATATAATTATTAT CAAATGTACATA
            || ||| |||||||| | ||    |   |||| |  |||     |  | | | | |
60          TTTTATGTATTTCAGAATTAACCTCCAAAAACATTTATTGACACACTACTACTCTTTCCG

191         AAAATAAAATAAATTTAAAAATATATACTTCAACAACAT CAAGCGCAAAATCCATAAAA
              |   |  | ||  ||   || |  |  | |   ||||   ||  ||  |   | | |
120         TATTGACTCTCAA CTAGTCATTTCAAAATAATTGACATGTCAGAAACATGAGTTACACAT

250         AGAATCATATGAAGCAAACAGAAACTTGCATGCTGCTGCCATACGTGTAAGCTTAAAGCA
              |  |||||   ||||  ||  ||  | |  ||  | |  | | | |   |  | |
179         GGTTGCATAT TTGCAAGTAG  ACGCGAAACTTGTCACTTCCTTTACATTTGAGTTTC

310         GACCTCAAAACCATGGGTAGCCACCCCATAAAAATGCCAGAAATGACTGGCCATTCCATA
              | | |  ||  ||  |  | | | |  |||    |  || | |  ||   |   | |||
234         CAACACCTAATCA CGACAACGATCATATAGCTCTCGCATGCA AACAAACATATGCATG

370         CA TCTTACACGTTCTCGACATGC ACTCCCCTAAATTTCCTATAAGAGCCCATTAC CA
             | ||||||||||   |  ||||| | || | |   |  |||||||   || |  || |
294         TATTCTTGCACGTGAACTCCATGCGAGTCTCTTTTCTCACCTATAAATACCAGCCACACC

427         TCATCTACTATATTCCATAAGAGCATCTTTCTGATATTCAGTAACAA
            |   | ||  | |||  |    | ||         | |   || || ||
354         TTCAGCGACATTCTTGACTCGAACCAAAACATACGCACATAGCAAAAA

Matches = 184     Mismatches = 208     Unmatched = 13
Length = 407      Matches/length =  45.7 percent
```

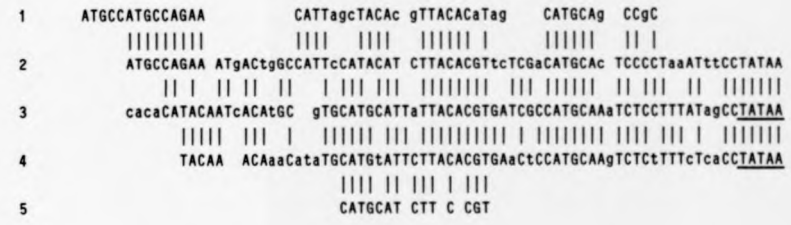
Figure 43

Comparison of homologous areas found in the regulatory regions of 2S albumin and ricin genes

Regions of homology found in the regulatory sequences of the genes encoding Brassica napus napin (napA), Arabidopsis thaliana 2S albumin (AT2S1), the castor bean preproalbumin (gcep6a) and the castor bean preproricin (gric) are indicated. A line between two adjacent sequences shows an identical base and a capital letter shows an identical base found in a non-adjacent sequence. Small letters indicate no homology. Gaps were introduced to increase homology. Numbering is as described by the authors. The 'TATA' boxes of napA and AT2S1 are underlined.

1	napA, bases 888 to 935
2	gcep6a, bases 343 to 414
3	napA, bases 1002 to 1073
4	AT2S1, bases -125 to -60
5	gric, bases 138 to 151

NapA gene sequence was taken from Josefsson et al (1987). AT2S1 gene sequence was taken from Krebbers et al (1988). Gric gene sequence taken from Tregear (1989).


1 ATGCCATGCCAGAA CATTagcTACAc gTTACACaTag CATGCAg CCgC
2 ATGCCAGAA ATgACtgGCCATTcCATACAT CTTACACGTtcTCGaCATGCAc TCCCCTaaATttCCTATAA
3 cacaCATACAATcACAtGC gTGCATGCATTaTTACACGTGATCGCCATGCAAaTCTCCTTTATagCCTATAA
4 TACAA ACAaaCataTGCATGtATTCTTACACGTGAaCtCCATGCAAgTCTCtTTTcTcaCCTATAA
5 CATGCAT CTT C CGT

D4.7 Homologies observed in the promoter regions of 2S albumin genes

Whist few homologies could be found when comparing the regulatory regions of the preproalbumin gene with the sunflower albumin gene (Allen et al, 1987), or the pea 2S albumin gene (Higgins et al, 1986), interesting homologies were noted upstream of the TATA boxes in the Brassica napin gene (Figure 41) and the Arabidopsis 2S albumin AT2S1 gene (Figure 42). These were noted to be the same regions and a comparison of all the homologies found (including that in the preproricin gene) is shown in Figure 43. This conserved region upstream of the regulatory TATA boxes strongly suggests involvement in temporal and tissue specific expression. This area includes the CATGCAT box described previously (D4.6) which is not entirely conserved in all the sequences, as well as an incomplete form of this motif (CATGCA) which is. The TATA boxes for the Arabidopsis 2S albumin and Brassica napin have been defined, and it suggested that the TATA box of the preproalbumin gene gcep6a occurs at the 3' end of this region also (bases 409-413).

The conserved region was found to be repeated twice in the Brassica napin clone. Parts of both of these regions overlapped with the repeating purine/pyrimidine regions noticed by Scofield and Crouch (1987), who speculated that the latter may be important in gene regulation. The repeating purine/pyrimidines have been seen in viral enhancers (Lusky et al, 1983) and are implicated in gene regulation.

In their genomic clone of the Brassica napin, Josefsson et al (1987) located a 5' hairpin loop and a TACACAT consensus repeat region. Speculation was made that these may be involved in transcriptional activation, for which there are many precedents in the literature: Degenerate (or non-degenerate) repeats as well as alterations in DNA topology (cruciform structures) have been implied in gene regulation in several systems (Gidoni et al. 1985; Hall et al. 1982; Harland et al. 1983; reviewed by Sarfling et al. 1985). The TACACAT consensus heptamer occurred twelve times and included the trimer CAC which featured eleven times in the promoter of the Brassica napin gene. Nine of these heptamers and six of the trimers occurred within the two sequence regions found in the Brassica napin that possessed homology to the other albumins (Figure 43).

Regulatory elements conferring endosperm specific expression have been located in storage protein genes from barley, wheat and maize (Forde et al. 1985). These conferred tissue specific expression in transformed tobacco (Colot et al. 1987). The consensus sequence TGTAAAG was not located by Scofield and Crouch (1987) in the Brassica napin, nor here in the preproalbumin gene.

Chen et al (1986) tried to define the regulatory elements that caused the expression of the alpha subunit of beta-conglycinin, a seed storage protein of soybean, to be both temporally and tissue specifically regulated in Petunia (Beachy et al. 1985). This was also demonstrated in tobacco later by Barker et al (1988). A consensus sequence of A(G/C/A)CCCA was suggested. This is not located within the regions of homology noted here (Figure 43), but is present twice in the

preproalbumin gene gcep6a at bases 332-337, which is 15 bases upstream of the conserved region of homology, and immediately downstream of the predicted TATA box (bases 415-420). Their locations, in bounding the conserved region, suggest that they may be significant, although their infrequency of appearance and their absence in the conserved region causes some doubt.

Other examples of cloned storage protein genes being correctly expressed developmentally in homologous and heterologous plant systems include the Brassica napin reintroduced into Brassica napus (Radke et al, 1988), and beta-phaseolin from Phaseolus vulgaris introduced into tobacco plants (Hoffmann et al, 1988; Greenwood and Chrispeels, 1985; Sengupta-Gopalan et al, 1985).

It is worth noting that whilst all of the homologies discovered in Figure 43 are from seed storage proteins (although the position of ricin as a storage protein is debatable), the tissue localisations of these are different. Ricin and the 2S albumins of castor bean are localised in the protein bodies of endosperm cells but not in the cotyledons (Kermode et al, 1985). The Brassica napins are found in the cotyledons and axis cells only (Crouch and Sussex, 1981). The Arabidopsis 2S albumins are only defined as located in the seed pods (Krebbers et al, 1988), but it is expected that they will be similar to the 12S seed storage protein which is transcribed only in the cotyledons and hypocotyl of the seed pod (Pang et al, 1988). Arabidopsis thaliana and Brassica napus are both members of the family Cruciferae, whilst Ricinus communis belongs to the family Euphorbiaceae (Youle and Huang, 1981). If the regions of homology defined in the promoter region are to include signals specifying gene

regulation in a tissue specific manner then similar trans-acting factors are presumably located only in the specific tissues of these plants.

It has been shown in other plant systems that the cis signals involved in regulating transcriptional initiation usually are located reasonably close to the transcribed part of the gene (Kamlen *et al*, 1986; Morelli *et al*, 1985). This would appear to be supported here by the close conservation of sequence between the clones gcep6a and gcep5l for 262 bases upstream of the translation start point, past which the sequences show little homology. This was also seen with the *Arabidopsis* 2S albumin gene family where all four members diverged rapidly from one another 210 bases upstream from the translational start point such that no homologies thereafter could be found. Such observations assume that the genes concerned are active.

Since the two napin genes and the *Arabidopsis* 2S albumin gene were all published simultaneously, no comparisons were made between the two sequences. With the arrival of the castor bean preproalbumin genes here, and the preproricin gene, homology studies have now been performed on the promoter elements upstream of the TATA box. Significant homologies have been found in one region only and it is here that further studies should be concentrated to determine what function these sequences play. This could be done by *in vitro* mutagenesis with subsequent transformation studies, and by defining the trans-acting factors by study of regulatory mutants and DNA binding proteins.

D4.8 Conclusions

The genes encoding the preproalbumin show temporally regulated expression in the seeds of castor bean. The preproalbumin is represented by a gene family of at least four members. Two of these members appear to have been cloned fully, whilst a third clone is incomplete. Sequence analysis of one showed it to be almost identical to the cDNA clones already isolated and sequenced. Comparison of the promoter regions of the two complete genes suggest that some of the information required for their regulation may be located within 250 bp of the translation start point. Comparison of the promoter regions with other 2S albumin gene promoters showed a highly conserved region of approximately 70 bp proceeding upstream from the TATA boxes. This region may define the sequences required for tissue and temporal specific regulation of these genes. More limited homology with the ricin gene promoter was seen also in this region.

D5 OVERALL SUMMARY OF RESULTS AND CONCLUSIONS

The precursor to the major castor bean 2S albumin has been cloned from a newly synthesised cDNA library and characterised. In many ways the precursor is similar to those of other 2S albumins. It possesses an N-terminal signal sequence that is cotranslationally removed, followed by post-translational endoproteolytic cleavage to liberate the mature subunits into the protein bodies. The mature protein located at the carboxyl terminus of the preproalbumin is believed to be the definitive sequence and corrects that obtained earlier by direct protein sequencing (Sharief and Li, 1982). The production of an 11 kDa mature protein from a 29 kDa precursor seemed more wasteful than processing events seen for all other 2S albumins. Closer examination of the precursor identified two areas of sequence that could correspond to the small and large subunits of a second 2S albumin.

Examination of the total soluble proteins from castor bean protein bodies showed that the 2S albumin fraction was composed of 5 major polypeptides, three of which were associated with the previously characterised 2S albumin. The other two polypeptides appeared reasonable candidates for the putative small and large subunits of the second 2S albumin predicted on the precursor. It seemed more than likely then, that the deposition of the two 2S albumins in the protein bodies of castor beans was mediated by a single precursor. The process would be much more efficient than if a different precursor was used for each (83% as opposed to 38% w/w). The size heterogeneity of the previously characterised 2S albumin appeared to be the result of the interaction of a single sized large subunit with one of two differently sized small subunits, probably generated by an endoproteolytic processing event.

Transcripts encoding the precursor were seen to be generated during the latter stages of seed development, indicating temporal regulation of the genes. The genes appeared to be grouped as a family of four members. Two of these and a third incomplete member were isolated from a previously generated castor bean genomic library. Sequence analysis of the upstream regions of the two complete genes suggested that the regulatory regions involved were conserved within a short stretch of sequence. Comparisons made with other 2S albumins genes showed a very highly conserved region existed for 70 bp immediately upstream of the TATA boxes. Within this region was found a small stretch of homology with the ricin gene, which is known to be regulated in a similar fashion to the castor bean 2S albumin. These areas are proposed to be strongly involved in the specific regulation of the preproalbumin genes and should be the subject of further experimentation to verify this.

REFERENCES

Abdullah R., Cocking E.C. and Thompson J.A. (1986) Biotechnology 4 1087-1090.

Allen R.D., Cohen E.A., Vonder Haar R.A., Adams C.A., Ma D.P., Nessler C.L. and Thomas T.L. (1987) Mol. Gen. Genet. 210 211-218.

Altenbach S.B., Pearson K.W., Leung F.W. and Sun S.S. (1986) in "Molecular Biology of Seed Storage Proteins and Lectins" Shannon L.M. and Chrispeels M.J. (Eds.). Americ. Soc. Plant. Physiol. pp137-146.

Altenbach S.B., Pearson K.W., Leung F.W. and Sun S.S. (1987) Plant Mol. Biol. 8 239-250.

Altschul A.M., Neucere N.J., Woodham A.A. and Dechary J.M. (1964) Nature 203 501-504.

Altschul A.M., Yatsu L.Y., Ory R.L. and Engleman E.M. (1966) Annu. Rev. Plant Physiol. 17 113-136.

Altschuler M. and Mascarenhas J.P. (1982) Plant Mol. Biol. 1 103-115.

Amersham International plc. (1985) "Membrane Transfer and Detection Methods" Buckinghamshire, UK.

Anderson S. and Kingston I.B. (1983) Proc. Natl. Acad. Sci. USA 80 6838-6842.

Aoyagi K., Kuhlemeier C. and Chua N-H. (1988) Mol. Gen. Genet. 213 179-185.

Banerji I., Olson L. and Schaffner W. (1983) Cell 33 729-740.

Barbieri L. and Stirpe F. (1982) Cancer Surveys 1 489-520.

Barker S.J., Harada J.J. and Goldberg R.B. (1988) Proc. Natl. Acad. Sci. USA 85 458-462.

Barnard J.H. (1930) J. Allergy 1 473-474.

Bassuner R., Huth A., Manteuffel R. and Rapoport T. (1983) Eur J. Biochem. 133 321-326.

Baumlein H., Wobus U., Pustell J. and Kafatos F.C. (1986) Nucleic Acids Research 14 2707-2719.

Bazett-Jones D.P., Yeckel M. and Gottesfeld J.M. (1985) Nature 317 824-828.

Beachy R.N., Chen Z.-L., Horsch R.B., Rogers S.G., Hoffmann N.J. and Fraley R.T. (1985) EMBO J. 4 3047-3053.

Beevers H. (1979) Annu. Rev. Plant Physiol. 30 159-197.

Bell J.N., Ryder T.B., Wingate V.P.M., Bailey J.A. and Lamb C.J. (1986) Mol. Cell. Biol. 6 1615-1623.

Bennett R.H. and Schwartz E. (1934) J. Allergy 5 427-431.

Bemoist C. and Chambon P. (1981) Nature 290 304-310.

Benoist C., O'Hare K., Breathnach R. and Chambon P. (1980) Nucleic Acids Research 8 127-142.

Bernton H.S. (1923) Amer. J. med Sci. 165 196.

Bernton H.S. (1945) Sth. Afr. med. J. 38 670-677.

Bewley J.D. and Black M. (1978) "Physiology and Biochemistry of Seeds in Relation to their Germination" Volume 1. Springer-Verlag. New York.

Bewley J.D. and Black M. (1985) "Seeds: Physiology of Development and Germination" Planum press. New York.

Birnstiel M., Busslinger M. and Strub K. (1985) Cell 41 349-359.

Bollini R., Vitale A. and Chrispeels M.J. (1983) J. Cell Biol. 96 999-1007.

Boshart M., Weber F., Jahn G., Dorsch-Hasler K., Fleckenstein B. and Schaffner W. (1985) Cell 41 521-530.

Boulter D. (1984) Philos. Trans. Roy. Soc. (Lond.) Ser. B. 304 323-332.

Bowles D.J., Marcus S.E., Pappin D.J.C., Findlay J.B.C., Elipolons E., Maycox P.R. and Burgess J. (1986) J. Cell Biol. 102 1284-1297.

Bray E. and Beachy R.N. (1985) Plant Physiol. 79 746-750.

Breathnach R. and Chambon P. (1981) Ann. Rev. Biochem. 50 349-383.

Briggs M.R., Kadonaga J.T., Bell S.P. and Tjian R. (1986) Science 234 47-52.

Broglie K.E., Gaynor J.J., and Broglie R.M. (1986) Proc. Natl. Acad. Sci. USA 83 6820-6824.

Buell G.N., Wickens M.P., Payvar F. and Schimke R.T. (1978) J. Biol. Chem. 253 2471-2482.

Burgess T.L. and Kelly R.B. (1987) Annu. Rev. Cell Biol. 3 243-293.

Butterworth A.G. and Lord J.M. (1983) Eur. J. Biochem. 137 57-65.

Campos F.A.P and Richardson M. (1983) FEBS Letts. 152 300-304.

Casey R., Domoney C. and Ellis N. (1986) Oxford Surveys of Plant Molecular and Cell Biology 3 1-95.

Cawley D.B. and Houston L.L. (1979) Biochim. et Biophys. Acta 581 51-62.

Chandler P.M., Spencer D., Randall P.J. and Higgins T.J.V. (1984) Plant Physiol. 75 651-657.

Chen Z.-L., Schuler M.A. and Beachy R.N. (1986) Proc. Natl. Acad. Sci. USA 83 8560-8564.

Chou P.Y. and Fasman G.D. (1978a) Annu. Rev. Biochem. 47 251-276.

Chou P.Y. and Fasman G.D. (1978b) Adv. Enzymol. Relat. Areas Mol. Biol. 47 45-148.

Chrispeels M.J. (1983) Planta 158 140-151.

Chrispeels M.J. (1984) Phil. Trans. R. Soc. Lond. Ser. B 304 309-322

Chrispeels M.J., Higgins T.J.V., Craig S. and Spencer D. (1982a) J. Cell Biol. 93 5-14.

Chrispeels M.J., Higgins T.J.V. and Spencer D. (1982b) J. Cell Biol. 93 306-313.

Chua K.Y., Stewart G.A., Thomas W.R., Simpson R.J., Dilworth R.J., Plozza T.M. and Turner K.J. (1988) J. Exp. Med. 167 175-182.

Colot V., Robert L.S., Kavanagh T.A., Bevan M.W. and Thompson R.D. (1987) EMBO Journal 6 3559-3564

Conner B.J., Reyes A.A., Morin C., Itakura K., Teplitz R.L. and Wallace R.B. (1983) Proc. Natl. Acad. Sci. USA 80 278-282.

Coombs R.R.A and Gell P.J.H. (1975) in "Clinical Aspects of Immunology" 3rd Edition, Gell P.J.H., Coombs R.R.A. and Lachmann P.J. (Eds.) Blackwell Scientific Publications, Oxford.

Cooper P., Ho T-H.D. and Hauptmann R.M. (1984) Plant Physiol. 75 431-441.

Creek K.E. and Sly W.S. (1984) in "Lysosomes in Pathology and Biology" Dingle J.T., Dean R.T. and Sly W.S. (Eds.) Elsevier Biomedical press. North Holland. pp63-82.

Crouch M.L. and Sussex I.M. (1981) Planta 153 64-74.

Crouch M.L., Tenbarge K.M., Simon A.E. and Feri R. (1983) J. Mol. Appl. Genet. 2 273-283.

Croy R.R.D., Gatehouse J.A., Evans D.M., and Boulter D. (1980) Planta 148 49-56.

Davies R.W., Botstein D. and Roth J.R. (1980) "Advanced Bacterial Genetics" Cold Spring Harbour, New York.

De Castro L.A.B., Lacerda Z., Aramayo R.A., Sampaio M.J.A.M. and Gander E.S. (1987) Mol. Gen. Genet. 206 338-343.

Della-Cioppa G., Kishore G.M., Beachy R.N. and Fraley R.T. (1987) Plant Physiol. 84 965-968.

Dickinson C.D., Evans R.P. and Nielson N.C. (1988) Nucleic Acids Research 16 371.

Dieckert J.W. and Dieckert M.C. (1976) J. Food Sci. 41 475-482.

Dieckert J.W. and Dieckert M.C. (1985) "The Chemistry and Biology of Seed Storage Proteins" in New Protein Foods Vol. 5: Seed Storage Proteins. Altschul A.M. and Wilcke H.L. (Eds.). Academic Press. London.

Dierks-Ventling C. and Ventling D. (1982) FEBS Letts. 142 147-150.

Diffley J.F.X. and Stillman B. (1986) Mol. Cell Biol. 6 1363-1373.

Dorel C., Voelker T.A., Herman E.M. and Chrispeels M.J. (1988) Plant Physiol. 86 (suppl.) 84.

Doyle J.J., Ladin B.F. and Beachy R.N. (1985) Biochem. Syst. Ecol. 13 123-132.

Doyle J.J., Schuler M.A., Godette W.D., Zenger V., Beachy R.N. and Slightom J.L. (1986) J. Biol. Chem. 261 9228-9238.

Dudman W.F. and Millerd A. (1975) Biochem. Syst. Ecol. 3 25-33.

Dynan W.S. and Tjian R. (1985) Nature 316 774-778.

Efstratiadis A., Kafatos F.C., Maxam A.M. and Maniatis T. (1976) Cell 7 279-288.

Efstratiadis A., Posakony J.W., Maniatis T., Lawn R.M., O'Connell C., Spritz R.A., DeRiel J.K., Forget B.G., Weissman S.M., Slightom J.L., Blechl A.E., Smithies O., Baralle F.E., Shoulders C.C. and Proudfoot N.J. (1980) Cell 21 653-668.

Endo Y., Mitsui K., Motizuki M. and Tsurugi K. (1987) J. Biol. Chem. 262 5908-5912.

Endow S.A. and Roberts R.J. (1977) J. Mol. Biol. 112 521-529.

Engleman E.M. (1966) Am. J. Bot. 53 231-237.

Ericson M.L., Rodin J., Lenman M., Glimelius K., Josefsson L.-G. and Rask, L. (1986) J. Biol. Chem. 261 14576-14581.

Fabijanski S. and Altosaar I. (1985) Plant Mol. Biol. 4 211-218.

Fang K.S.Y., Vitale M., Fehlner P. and King T.P. (1988) Proc. Natl. Acad. Sci. USA 85 895-899.

Feinberg A.P. and Vogelstein B. (1983) Analytical Biochem. 132 6-13.

Festenstein G.N., Hay F.C., Miflin B.J. and Shewry P.R. (1984) Planta 162 524-531.

Figley K.D. and Elrod R.H. (1928) J. Amer. Med. Ass. 90 79.

Follweiler F.L. and Haley D.E. (1925) J. Amer. med. Ass. 84 1418.

Forde B.G., Heyworth A., Pywell J. and Kreis M. (1985) Nucleic Acids Research 13 7327-7339.

Forde J., Forde B.G., Fry R.P., Kreis M., Shewry P.R. and Miflin B.J. (1983) FEBS Letts. 162 360-366.

Fowden L. and Miflin B.J. (Eds.) (1984) "Seed storage proteins" Phil. Trans. R. Soc. Lond. B 304 273-407.

Franz H. (1988) Lectins: Biol., Biochem., Clin. Biochem. 6 7-13.

Fried M. and Crothers D.M. (1981) Nucleic Acids Research 9 6505-6525.

Fuchs L.Y., Covarrubias L., Escalante L., Sanchez S. and Bolivar F. (1980) Gene 10 39-46.

Fukasawa T., Hara-Nishimura I. and Nishimura M. (1988) Plant Cell Physiol. 29 339-345.

Galas D.J. and Schmitz A. (1978) Nucleic Acids Research 5 3157-3170.

Garcia-Olmedo F., Salcedo G., Sanchez-Monge R., Gomez L., Royo J. and Carbonero P. (1987) Oxford Surveys of Plant Molecular and Cell Biology 4 275-334.

Garver W.P. (1948) J. Lab. clin. Med. 33 1613.

Gatehouse J.A., Croy R.R.D. and Boulter D. (1980) Biochem J. 185 497-506.

Gatehouse J.A., Evans I.M., Croy R.R.D., and Boulter D. (1986) Philos. Trans. R. Soc. Lond. Ser. B. 314 367-384.

Gatehouse J.A., Brown D., Gilroy J., Levasseur M., Castleton J. and Ellis T.H.N. (1988) Biochem J. 250 15-24.

Gennis L.S. and Contor C.R. (1976) J. Biol. Chem. 251 741-746.

Geraghty D., Peifer M.A., Rubenstein I. and Messing J. (1981) Nucleic Acids Research 9 5163-5174.

Ghosh P.K., Lebowitz P., Frisque R.J. and Gluzman Y. (1981) Proc. Natl. Acad. Sci. USA 78 100-104.

Gidoni D., Kadonaga J.T., Barrera-Saldana H., Takahashi K., Chambon P. and Tjian R. (1985) Science 230 511-517.

Gifford D.J. and Bewley J.D. (1983) Plant Pyhsiol. 72 376-381.

Gifford D.J. and Bewley J.D. (1984a) Plant Pyhsiol. 74 1006-1009.

Gifford D.J. and Bewley J.D. (1984b) Plant Science Letters 36 37-42.

Gifford D.J., Greenwood J.S. and Bewley J.D. (1982) Plant Physiol. 69 1471-1478.

Gifford D.J., Greenwood J.S. and Bewley J.D. (1983) J. Exp. Bot. 34 1433-1443.

Gingeras T.R., Greenough L., Schildkraut I. and Roberts R.J. (1981) Nucleic Acids Research 9 4525-4536.

Glover D.M. (Ed.) (1985) "DNA Cloning Volume 1: A Practical Approach." IRL Press, Oxford.

Gorski K., Carneiro M. and Schibler U. (1986) Cell 47 767-776.

Graham J.S., Jennings A.C., Morton R.K., Palk B.A. and Raison J.K. (1962) Nature 196 967-969.

Graham J.S., Pearce G., Merryweather J., Titani K., Ericsoon L. and Ryan C.A. (1985) J. Biol. Chem. 260 6555-6560.

Greenwood J.S. (1983) Ph.D. Thesis. University of Calgary. Alberta.

Greenwood J.S. and Bewley J.D. (1982) Can. J. Bot. 60 1751-1760.

Greenwood J.S. and Chrispeels M.J. (1985a) Plant Physiol. 79 65-71.

Greenwood J.S. and Chrispeels M.J. (1985b) Planta 164 295-302.

Grosschedl R. and Birnstiel M.L. (1980) Proc. Natl. Acad. Sci. USA 77 1432-1436.

Grosschedl R., Wasylyk B., Chambon P. and Birnstiel M.L. (1981) Nature 294 178-180.

Grunstein M. and Hogness D.S. (1975) Proc. Natl. Acad. Sci. USA 72 3961-3965.

Gruss P. (1984) DNA 3 1-5.

Gubler U. and Hoffman B.J. (1983) Gene 25 263-269.

Guldager P. (1978) Theor. Appl. Genet. 53 241-250.

Halford N.G., Morris N.A., Urwin P., Williamson M.S., Kasarda D.D., Lew E.J-L., Kreis M. and Shewry P.R. (1988) Biochim. et Biophys. Acta 950 435-440.

Hall B.D., Clarkson S.G. and Tocchini-Valentini G. (1982) Cell 29 3-5.

Halling K.C., Halling A.C., Murray E.E., Ladin B.F., Houston L.L. and Weaver R.F. (1985) Nucleic Acids Research 13 8019-8033.

Halpin C., Conder M.J. and Lord J.M. (1989) Planta (in press).

Hanahan D. (1985) in "DNA cloning Volume 1: a practical approach" Glover D.M. (Ed.) IRL Press. Oxford.

Hara-Nishimura I. and Nishimura M. (1987) Plant Physiol. 85 440-445.

Harland R.M., Weintraub H. and McKnight S.L. (1983) Nature 302 38-43.

Harley S.M. and Lord J.M. (1985) Plant Science 41 111-116.

Hart G.W., Brew K., Gregory A.G., Bradshaw R.A. and Lennarz W.J. (1979) J. Biol. Chem. 254 9747-9753.

Hartig T. (1855) Bot. Ztg. 13 881-882.

Hartig T. (1856) Bot. Ztg. 14 257-268.

Hattori T., Ichihara S. and Nakamura K. (1987) Eur. J. Biochem. 166 533-538.

Heidecker G. and Messing J. (1986) Ann. Rev. Plant Physiol. 37 439-466.

Hemperly J.J., Mostov K.E. and Cunningham B.A. (1982) J. Biol. Chem. 257 7903-7909.

Hentschel C.C. and Birnstiel M.L. (1981) Cell 25 301-313.

Herman E.M., Shannon L.M. and Chrispeels M.J. (1986) in "Molecular Biology of Seed Storage Proteins and Lectins" Shannon L.M. and Chrispeels M.J. (Eds.) The American Society for Plant Physiologists. pp163-174

Herman E.M., Chrispeels M.J. and Hoffman L.M. (1989) Cell Biol. Int. Rep. 13 37-45.

Higgins T.J.V and Spencer D. (1981) Plant Physiol. 67 205-211.

Higgins T.J.V. (1984) Ann. Rev. Plant Physiol. 35 191-221.

Higgins T.J.V., Chandler P.M., Zurawski G., Buttons S.C. and Spencer D. (1983a) J. Biol. Chem. 258 9544-9549.

Higgins T.J.V., Chrispeels M.J., Chandler P.M. and Spencer D. (1983b) J. Biol. Chem. 258 9550-9552.

Higgins T.J.V, Chandler P.M., Randall P.J., Spencer D., Beach L.R., Blagrove R.J., Kortt A.A. and Inglis A.S. (1986) J. Biol. Chem. 261 11124-11130.

Hilder V.A., Gatehouse A.M.R., Sheerman S.E., Barker R.F. and Boulter D. (1987) Nature 300 160-163.

Hoffman L.M. (1984) J. Mol. Appl. Genet. 2 447-453.

Hoffman L.M. and Donaldson D.D. (1985) EMBO Journal 4 883-889.

Hoffman L.M., Donaldson D.D. and Herman E.M. (1988) Plant Mol. Biol. 11 717-729.

Holmes D.S. and Quigley M. (1981) Analytical Biochem. 14 193-197.

Hopp T.P and Woods K.R. (1981) Proc. Natl. Acad. Sci. USA 78 3824-3828.

Huang D.L.R., Lin K.T.D., Yang W.K. and Foard D.E. (1977) Biochim. et Biophys. Acta 495 369-382.

Hurkman W.J., Smith L.D., Richter J. and Larkins B.A. (1981) J. Cell Biol. 89 292-299.

Hyldig-Nielsen J.J., Jensen E.O., Paludan K., Wiborg O., Garret R., Jorgensen P. and Marcker K.A. (1982) Nucleic Acids Research 10 689-701.

Ishizaka K. and Ishizaka T. (1967) J. Immunol. 99 1187-1198.

Jameson B.A. and Wolf H. (1988) Computer Applications in Biosciences 4 181-186.

Janzen D. (1983) in "Physiology Plant Ecology: Vol. 3" Lange O. (Ed.) Encyclopaedia of Plant Physiology, New Series, Vol. 12 pp625-656. Springer-Verlag.

Jaye M., DelaSalle H., Schamber F., Balland A., Kohli V., Findeli A., Tolstoshev P. and Lecocq J.P. (1983) Nucleic Acids Research 11 2325-2335.

Johnson L.M., Bankaitis V.A. and Emr S.D. (1987) Cell 48 875-885.

Jones K.A., Kadonaga J.T., Rosenfeld P.J., Kelly T.J. and Tjian R. (1987) Cell 48 79-89.

Josefsson L-G., Lenman M., Ericson M.L. and Rask L. (1987) J. Biol. Chem. 262 12196-12201.

Kamlen H., Schell J. and Kreuzaler F. (1986) EMBO J. 5 1-8.

Karin M., Haslinger A., Holtgreve H., Cathala G., Slater E. and Baxter J.D. (1984) Cell 36 371-379.

Kasarda D.D., Okita R.W., Bernardin J.E., Baecker P.A., Nimmo C.C., Lew E.J-L., Dietler M.D. and Greene F.C. (1984) Proc. Natl. Acad. Sci. USA 81 4712-4716.

Kashlan N. and Richardson M. (1981) Phytochemistry 20 1781-1784.

Kaulen H., Schell J. and Kreuzaler F. (1986) EMBO Journal 5 1-8.

Kelly R.B. (1985) Science 230 25-32.

Kemeny D.M., Frankland A.W., Fahkri Z.I. and Trull A.K. (1981) Clinical Allergy 11 463-471.

Kermode A.R. and Bewley J.D. (1985a) J. Exp. Bot. 36 1906-1915.

Kermode A.R. and Bewley J.D. (1985b) J. Exp. Bot. 36 1916-1927.

Kermode A.R. and Bewley J.D. (1986) J. Exp. Bot. 37 1887-1898.

Kermode A.R. and Bewley J.D. (1988) J. Exp. Bot. 39 487-497.

Kermode A.R., Gifford D.J., Thakore E. and Bewley J.D. (1985) J. Exp. Bot. 36 792-799.

Kilpatrick M.W., Klysik J., Singleton C.K., Zanling D.A., Jovin T.W., Hanan L.H., Erlanger B.F. and Wells R.D. (1984) J. Biol. Chem. 259 7268-7274.

Kirihara J.A., Hunsperger J.P., Mahoney W.C. and Messing J.W. (1988) Mol. Gen. Genet. 211 477-484.

Kitamura K., Takagi T. and Shibasaki K. (1976) Agric. Biol. Chem. 40 1837-1844.

Konarev V.G., Gavriljuk I.P. and Gubareva N.K. (1987) Seed Sci. and Technol. 15 675-678.

Kowles R.V. and Phillips R.I. (1985) Proc. Natl. Acad. Sci. USA 82 7010-7014.

Krebbers E., Herdies L., de Clerq A., Seurinck J., Leemans J., van Damme J., Segura M., Gheysen G., van Montagu M. and Vandekerckhove J. (1988) Plant Physiol. 87 859-866.

Kreis M., Forde B.G., Rahman S., Miflin B.J. and Shewry P.R. (1985a) J. Mol. Biol. 183 499-502.

Kreis M., Shewry P.R., Forde B.G., Forde J. and Miflin B.J. (1985b) Oxford Surveys of Plant Molecular and Cell Biology 2 253-317.

Kuhlemeier C., Green P.J. and Chua N-H. (1987) Ann. Rev. Plant Physiol. 38 221-257

Kyte J. and Doolittle R.F. (1982) J. Mol. Biol. 157 105-132.

Laemmli U.K. (1970) Nature 227 680-685.

Lamb F.I. (1984) Ph.D. Thesis, University of Warwick

Lamb F.I., Roberts L.M. and Lord J.M. (1985) Eur. J. Biochem. 148 265-270.

Laroche-Raynal M. and Delseny M. (1986) Eur. J. Biochem. 157 321-327.

Lathe R. (1985) J. Mol. Biol. 183 1-12.

Lazaro A., Rodriguez-Palenzuela P., Marana C., Carbonero P. and Garcia-Olmedo F. (1988) FEBS Letts. 239 147-150.

Lee W., Haslinger A., Karin M. and Tjian R. (1987) Nature 325 368-372.

Li S.S.-L., Lin T.T.-S and Forde M.D. (1977) Biochim. et Biophys. Acta 492 364-369.

Lilley G.G. and Inglis A.S. (1986) FEBS Letts. 195 235-241.

Lockey S.D. Jr. and Dunkelberger L. (1968) J. Amer. Med. Assoc. 206 2900.

Loenen W.A.M. and Blattner F.R. (1983) Gene 26 171-179.

Lonnerdahl B. and Janson J-C. (1972) Biochim. et Biophys. Acta 278 175-183.

Lord J.M. (1985a) Eur. J. Biochem. 146 403-409.

Lord J.M. (1985b) Eur. J. Biochem. 146 411-416.

Lord J.M. (1985c) New Phytol. 101 351-366.

Lord J.M., Lamb F.I. and Roberts L.M. (1984) Oxford Surveys of Plant Molecular and Cell Biology 1 85-101.

Lord J.M. and Robinson C.R. (1986) in "Plant Proteolytic Enzymes" (Dalling M.J., Ed.) pp69-80, CRC Press Inc., Boca Raton, FL.

Lord J.M., Halpin C., Conder M.J. and Irwin S.D. (1989) in "Compartmentation of Metabolism in Non-Photosynthetic Tissues" Emes M.J. (Ed.) Society for Experimental Biology Symposium Series. University Press. Cambridge.

Lord J.M., Spooner R.A., Hussain K. and Roberts L.M. (1988) Adv. Drug Delivery Rev. 2 297-318.

Lott J.N.A. (1980) in "The Biochemistry of Plants: A Comprehensive Treatise" Vol. 1. Tolbert N.E. (Ed.) Academic Press. New York. pp589-623.

Lusky M., Berg L., Weiher H. and Botchan M. (1983) Mol. Cell. Biol. 3 1108-1122.

Lycett G.W., Delauney A.J. and Croy R.R.D. (1983) FEBS Letts. 153 43-46.

Lycett G.W., Croy R.R.D., Shirsat A.H., Richards D.M. and Boulter D. (1985) Nucleic Acids Research 13 6733-6743.

Maeda K., Hase T. and Matsubara H. (1983) Biochim. et Biophys. Acta 743 52-57.

Mahoney W.C., Hermodson M.A., Jones B., Powers D.D., Corfman R.S. and Reeck G.R. (1984) J. Biol. Chem. 259 8412-8416.

Maier U.-G., Brown J.W.S., Toloczyki C. and Feix G. (1987) EMBO Journal 6 17-22.

Maniatis T., Jeffrey A. and Kleid D.G. (1975) Proc. Natl. Acad. Sci. USA 72 1184-1188.

Maniatis T., Fritsch E.F. and Sambrook J. (1982) "Molecular Cloning: A Laboratory Manual" Cold Spring Harbor Laboratory.

Mantell S.H., Matthews J.A. and McKee R.A. (1985) "Principles of Plant Biotechnology: An Introduction to Genetic Engineering in Plants" Blackwell Scientific Publications. Oxford.

Marcu K. and Dudock B. (1974) Nucleic Acids Research 1 1385-1397.

Marsh D.G., Goodfriend L., King T.P., Lowenstein H. and Platts-Mills T.A.E. (1988) Int. Arch. Allergy appl. Immun. 85 194-200.

Mason P.J., Jones M.B., Elkington J.A. and Williams J.G. (1985) EMBO J. 4 205-211.

Masumura T., Shibata D., Hibino T., Kato T., Kawabe K., Takeba G., Tanaka K. and Fujii S. (1989) Plant Mol. Biol. 12 123-130.

Matta N., Gatehouse J.A. and Boulter D. (1981) J. Exp. Bot. 32 183-197.

McGurl B. (1986) PhD Thesis, Warwick University.

Mendes E. and Ulhoa-Cintra A.C. (1954) J. Allergy 25 253.

Messing J., Geraghty D., Heidecker G., Hu N-T., Kridl J. and Rubenstein I. (1983) in "Genetic Engineering of Plants" Kosuge T., Meredith C.R. and Hollaender A. (Eds.) Plenum Press, New York. pp211-227.

Mizushima-Sugano J. and Roeder R.G. (1986) Proc. Natl. Acad. Sci. USA 83 8511-8515.

Montgomery D.L., Hall B.D., Gillam S. and Smith M. (1978) Cell 14 673-680.

Morelli G., Nagy F., Fraley R.T., Rogers S.G. and Chua N.-H. (1985) Nature 315 200-204.

Mosthaf L., Pawlita M. and Gruss P. (1985) Nature 315 597-600.

Mottier D.M. (1921) Ann. Bot. Lond. 35 350-364.

Muller M.M., Gerster T. and Schaffner W. (1988) Eur. J. Biochem. 176 485-495.

Murphy D.J., Cummins I. and Kang A.S. (1989) Biochem. J. 258 285-293.

Myers R.M., Tilly K. and Maniatis T. (1986) Science 232 613-618.

Nielson N.C. (1984) Philos. Trans. Roy. Soc. B. 304 287-296.

Nordheim A. and Rich A. (1983) Proc. Natl. Acad. Sci. USA 80 1821-1825.

Ochoa S. and deHaro C. (1979) Ann. Rev. Biochem. 48 549-580.

Odani S. and Ikenaka T. (1973) J. Biochem. 74 697-715.

Odani S. and Ikenaka T. (1978) J. Biochem (Tokyo) 83 737-745.

Odani S., Koide T. and Ono T. (1983a) J. Biol. Chem. 258 7998-8003.

Odani S., Koide T., Ono T. and Ohnishi K. (1983b) Biochem J. 213 543-545.

Odani S., Koide T. and Ono T. (1983c) J. Biochem. 93 1701-1704.

Okayama H. and Berg P. (1982) Mol. Cell. Biol. 2 161-170.

Old R., Murray K. and Roizes G. (1975) J. Mol. Biol. 92 331-339.

Ordman D. (1955) Int. Arch. Allergy 7 10-24.

Olsnes S. and Pihl A. (1982) in "Molecular Action of Toxins and Viruses" Cohen P. and van Heyningen S. (Eds.) Elsevier. New York. pp51-105.

Olsnes S., Saltvedt E. and Pihl A. (1974) J. Biol. Chem. 249 803-810.

Osborne T.B., Mendel L.B. and Harris I.F. (1905) Amer. J. Physiol. 14 259.

Osborne T.B. (1924) "The Vegetable Proteins" 2nd ed. Longman, Green and Co. London.

Pang P.P., Pruitt R.E. and Meyerowitz E.M. (1988) Plant Mol. Biol. 11 805-820.

Panzani R. (1957) Int. Arch. Allergy 11 224-236.

Panzani R. and Layton L.L. (1963) Int. Arch. Allergy 22 350-368.

Pappenheimer A.M., Olsnes S. and Harper A.A. (1974) J. Immunol. 113 835-841.

Pederson K., Devereux J., Wilson D.R., Sheldon E. and Larkins B.A. (1982) Cell 29 1015-1026.

Pelham H.R.B. and Jackson R.J. (1976) Eur. J. Biochem. 67 247-257.

Pfeffer S.R. and Rothman J.E. (1987) Annu. Rev. Biochem. 56 829-852.

Pfeffer W. (1872) Jahrb. Wiss. Bot. 8 429-574.

Pless D.D. and Lennarz W.J. (1977) Proc. Natl. Acad. Sci. USA 74 134-138.

Proudfoot N.J. and Brownlee G.G. (1976) Nature 263 211-214.

Quatrano R.S. (1986) Oxford Surveys of Plant Molecular and Cell Biology 3 467-477.

Radke S.E., Andrews B.M., Moloney M.M., Crouch M.L., Kridl J.C. and Knauf V.C. (1988) Theor. Appl. Genet. 75 685-694.

Renan M.J. (1987) Gene 60 245-254.

Rhodes C.A., Lowe K.S. and Ruby K.L. (1988) Biotechnology 6 56-60.

Robbins W.J.P. (1923) Science 58 305-306.

Robert L.S., Nozzolillo C. and Altosaar I. (1985) Biochem J. 226 847-852.

Roberts L.M., Lamb F.I. and Lord J.M. (1985) J. Biol. Chem. 260 15682-15686.

Roberts L.M., Lamb F.I. and Lord J.M. (1987) in "Membrane-mediated Cytotoxicity" Bonavida B. and Collier R.J. (Eds.) Alan R. Liss. New York. pp73-82.

Roberts L.M. and Lord J.M. (1981a) Planta 152 420-427.

Roberts L.M. and Lord J.M. (1981b) Eur. J. Biochem. 119 31-41.

Roberts R.J. (1983) Nucleic Acids Research 11 r135-r167.

Roitt I.M., Brostoff J. and Male D.K. (1985) "Immunology" Gower Medical publishing. London.

Roussell D.L., Boston R.S., Goldsbrough P.B. and Larkins B.A. (1988) Mol. Gen. Genet. 211 202-209.

Rusche J.R. and Howard-Flanders P. (1985) Nucleic Acids Research 13 1997-2008.

Ryan C.A. (1973) Annu. Rev. Plant Physiol. 24 173-196.

Ryan C.A. (1981) "Proteinase Inhibitors" in "The Biochemistry of Plants" Vol. 6. Academic Press, New York. pp. 351-370.

Sanger F., Nicklen S. and Coulson A.R.. (1977) Proc. Natl. Acad. Sci. USA 74 5463-5467.

Schibler U. and Sierra F. (1987) Ann. Rev. Genet. 21 237-257.

Schoffl F., Baumann G., Raschke E. and Bevan M. (1986) Philos. Trans. R. Soc. Lond. Ser. B. 314 453-468.

Scholer H.R. and Gruss P. (1984) Cell 36 403-411.

Scholer H.R. and Gruss P. (1985) EMBO Journal 4 3005-3013.

Schulze-Lefert P., Dangl J.L., Becker-Andre M., Hahlbrock K. and Schulz W. (1989) EMBO Journal 8 651-656.

Scofield S.R. and Crouch M.L. (1987) J. Biol. Chem. 262 12202-12208.

Sengupta-Gopalan C., Reichert N.A., Barker R.F., Hall T.C. and Kemp J.D. (1985) Proc. Natl. Acad. Sci. USA 82 3320-3324.

Serfling E., Jasin M. and Schaffner W. (1985) Trends Genet. 1 224-230.

Shagger H. and von Jagow G. (1987) Analytical Biochemistry 166 368-379.

Sharief F.S. and Li S.S-L. (1982) J. Biol. Chem. 257 14753-14759.

Shewry P.R. (1988) Plants Today 1 147-151.

Shewry P.R., Lafiandra D., Salcedo G., Aragoncillo C., Garcia-Olmedo F., Lew E.J-L., Dietler M.D. and Kasarda D.D. (1984a) FEBS Letts. 175 359-363.

Shewry P.R., Miflin B.J. and Kasarda D.D. (1984b) Phil. Trans. R. Soc. Lond. B. 304 297-308.

Shishodo K. and Ando T. (1981) Biochim. et Biophys. Acta 656 123-127.

Slightom J.L. and Chee P.P. (1987) "Advances in the Molecular Biology of Plant Seed Storage Proteins" in Biotechnology Advances Vol. 5. Pt. 1. pp29-45. Pergamon Journals Ltd.

Slightom J.L., Sun S.M. and Hall T.C. (1983) Proc. Natl. Acad. Sci. USA 80 1897-1901.

Snell W.H. (1924) Science 59 300.

Snell M.A. (1952) Arch. industr. Hyg. occup. Med. 6 113-115.

Sodroski J.G., Rosen C.A. and Haseltin W.A. (1984) Science 225 381-384.

Southern E.M. (1975) J. Mol. Biol. 98 503-517.

Spencer D. (1984) Phil. Trans. R. Soc. Lond. B 304 275-285.

Spencer D. and Higgins T.J.V. (1980) Biochem. Int. 1 502-509.

Spencer D., Chandler P.M., Higgins T.J.V, Inglis A.S., and Rubira M. (1983) Plant Mol. Biol. 2 259-267.

Spies J.R. and Coulson E.J. (1943) J. Amer. Chem. Soc. 65 1720-1728.

Stillmark H. (1888) Inaugural Dissertation, University of Dorpat, Estonia.

Stinissen H.M., Peumans W.J. and Carlier A.R. (1983) Plant Mol. Biol. 2 33-40

Sumner J.B. (1926) J. Biol. Chem. 69 435.

Sun S.S.M, Altenbach S.B. and Leung F.W. (1987) Eur. J. Biochem. 162 477-483.

Tague B.W. and Chrispeels M.J. (1987) J. Cell Biol. 105 1971-1979.

Tague B.W. and Chrispeels M.J. (1988) Plant Physiol. 86 (suppl.) 84.

Takahashi Y., Kato K., Hayashizaki Y., Wakabayashi T., Ohtsuka E., Matsuki S., Ikehara M. and Matsubara K. (1985) Proc. Natl. Acad. Sci. USA 82 1931-1935.

Templeman T.S., Demaggio A.E. and Stein D.B. (1986) Am. J. Bot. 73 688.

Thomas P. (1980) Proc. Natl. Acad. Sci. USA 77 5201-5205.

Thomas W.R., Stewart G.A., Simpson R.J., Chua K.Y., Plozza T.M., Dilworth R.J., Nisbet A. and Turner K.J. (1988) Int. Arch. Allergy appl. Immun. 85 127-129.

Thorpe S.C., Kemeny D.M., Panzani R.C., McGurl B. and Lord J.M. (1988) J. Allergy Clin. Immunol. 82 67-72.

Timko M.P., Kausch A.P., Castresana C., Fassler J., Herrera-Estrella L., Van den Broeck G., Van Montagu M., Schell J. and Cashmore A.R. (1985) Nature 318 579-582.

Tobin E.M. and Silverthorne J. (1985) Ann. Rev. Plant Physiol. 36 569-593.

Tosi M., Young R.A., Hagenbuchle O. and Schibler V. (1981) Nucleic Acids Research 9 2313-2323.

Towbin H., Staehlin T. and Gordon J. (1979) Proc. Natl. Acad. Sci. USA 76 4350-4354.

Tregear J. (1989) Ph.D. Thesis, University of Warwick.

Tully R.E. and Beevers H. (1976) Plant Physiol. 58 710-716.

Ullrich A., Berman C.H., Dull T.J., Gray A. and Lee J.M. (1984) EMBO J. 3 361-364.

United Nations Industrial Development Organisation (1974) "Castor Oil Production and Processing" United Nations. New York.

United States Biochemical Corporation (1988) "Sequenase: Step by Step Protocols for DNA Sequencing with Sequenase" 4th Edition. Ohio.

Valls L.A., Hunter C.P., Rothman J.H. and Stevens T.H. (1987) Cell 48 887-897.

van Deurs B., Ryde-Petersen L., Sundan A., Olsnes S. and Sandvig K. (1985) Exp. Cell Research 159 287-305.

van Deurs B., Tonnessen T.I., Petersen O.W., Sandvig K. and Olsnes S. (1986) J. Cell Biol. 102 37-45.

Verma I.M., Temple G.F., Fan H. and Baltimore D. (1972) Nature (New Biology) 235 163-167.

Vitale A. and Bollini R. (1986) in "Molecular Biology of Seed Storage Proteins and Lectins" Shannon L.M. and Chrispeels M.J. (Eds.) The American Society for Plant Physiologists. pp175-182

Vodkin L.O., Rhodes P.R. and Goldberg R.B. (1983) Cell 34 1023-1031.

Voelker T.A., Herman E.M. and Chrispeels M.J. (1989) The Plant Cell 1 95-104.

Voelker T.A., Staswick P. and Chrispeels M.J. (1986) EMBO Journal 5 3075-3082.

von Heijne G. (1983) Eur. J. Biochem. 133 17-21.

von Heijne G. (1984) J. Mol. Biol. 173 243-251.

von Heijne G. (1985) J. Mol. Biol. 184 99-105.

von Pirquet C. (1906) Muenchen Med. Wochenschr 53 1457.

Wallace J.C., Galili G., Kawata E.E., Cuellar R.E., Shotwell M.A. and Larkins B.A. (1988) Science 240 662-664.

Wallace R.B., Johnson M.J., Hirose T., Miyake T., Kawashima E.H. and Itakura K. (1981) Nucleic Acids Research 9 879-894.

Walling L., Drews G.N. and Goldberg R.B. (1986) Proc. Natl. Acad. Sci. USA 83 2123-2127.

Wasylyk B. (1988) Biochim. et Biophys. Acta 951 17-35.

Wasylyk B., Wasylyk C., Augereau P. and Chambon P. (1983) Cell 32 503-514.

Watson M.E.E. (1984) Nucleic Acids Research 12 5145-5164.

Weiher H., Konig M. and Gruss P. (1983) Science 219 626-631.

Weising K., Schell J. and Kahl G. (1988) Annu. Rev. Genet. 22 421-477.

Whitelaw E. and Proudfoot N. (1986) EMBO J. 5 2915-2922.

Wickens M.P., Buell G.N. and Schimke R.T. (1978) J. Biol. Chem. 253 2483-2495.

Wiederrecht G., Shuey D.J., Kibbe W.A. and Parker C.S. (1987) Cell 48 507-515.

Wieslander L. (1979) Analytical Biochem. 98 305-309.

Wilson G.A. and Young F.E. (1975) J. Mol. Biol. 97 123-125.

Woringer P. (1944) Z. klin. Med. 143 499-509. (abstracted in Biol. Abstr. No. 8567, in 1946).

Yanisch-Perron C., Vieira J. and Messing J. (1985) Gene 33 103-119.

Yoshimori R., Roulland-Dussoix D. and Boyer H.W. (1972) J. Bacteriol. 112 1275-1279.

Youle R.J. and Huang A.H.C. (1976) Plant Physiol. 58 703-709.

Youle R.J. and Huang A.H.C. (1978a) Plant Physiol. 61 13-16.

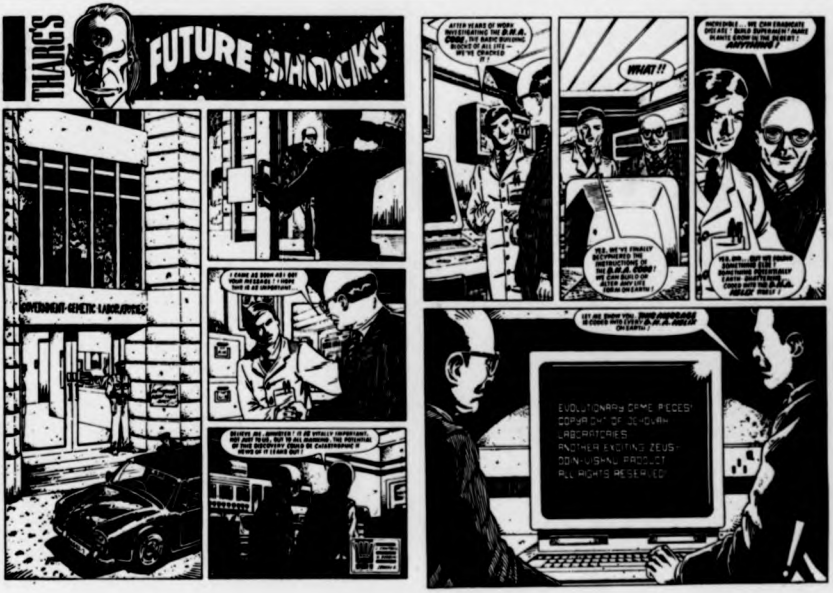
Youle R.J. and Huang A.H.C. (1978b) Plant Physiol. 61 1040-1042.

Youle R.J. and Huang A.H.C. (1979) J. Agric. Food Chem. 27 500-503.

Youle R.J. and Huang A.H.C. (1981) Amer. J. Bot. 68 44-48.

Originally appeared in "2000 AD" issue 604 (10/12/88)

Permission sought.


THARG'S
FUTURE SHOCKS
GOVERNMENT GENETIC LABORATORIES
AFTER YEARS OF WORK INVESTIGATING THE D.N.A. CODE, THE BASIC BUILDING BLOCKS OF ALL LIFE — WE'VE CRACKED IT!
WHAT!!
INCREDIBLE... WE CAN ERADICATE DISEASE! BUILD SUPERMEN! MAKE PLANTS GROW ON THE MOON! ANYTHING!
YES, WE'VE FINALLY DISCOVERED THE INSTRUCTIONS IN THE D.N.A. CODE! WE CAN BUILD OR ALTER ANY LIFE-FORM ON EARTH!
YES, UM... BUT WE SHOULD SOMETHING ELSE! SOMETHING POTENTIALLY EVEN MORE IMPORTANT! COMES INTO THE D.N.A. AREA! WHAT WAS IT?
I CAME AS SOON AS I GOT YOUR MESSAGE! I KNOW THIS IS AN INTERRUPTION!
BELIEVE ME, DOCTOR! IT IS VITALLY IMPORTANT, NOT JUST TO US, BUT TO ALL MANKIND, THE POTENTIAL OF THIS DISCOVERY COULD BE CATASTROPHIC IF MISUSED! IT MUST BE DESTROYED!
LET ME SHOW YOU SOMETHING ELSE COMES INTO THE D.N.A. AREA ON EARTH!
EVOLUTIONARY CRIME PIECES!
COPYRIGHT OF SCHOLAN LABORATORIES
ANOTHER EXCITING ZEUS CONSUMER PRODUCT
ALL RIGHTS RESERVED!

TITLE MOLECULAR CLONING OF THE MAJOR CASTOR BEAN 2S ALBUMIN PRECURSOR

AUTHOR STEPHEN D. IRWIN B.Sc.

INSTITUTION and DATE UNIVERSITY OF WARWICK 1989